

# **Evaluation eines Programms zur Motivationsförderung in der Schule**

Stärkung von Selbstwirksamkeitserwartung und Reduktion von Prüfungsängstlichkeit und  
Stresserleben bei Schülerinnen und Schülern

**Dissertation**  
**zur Erlangung des akademischen Grades**  
**Dr. phil.**  
**im Fach Erziehungswissenschaften**

eingereicht am 24. Februar 2015

verteidigt am 16. Juni 2015

an der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät  
der Humboldt-Universität zu Berlin

von

Dipl.-Psych. Katharina Chwallek

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekanin der Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät

Prof. Dr. Julia von Blumenthal

Gutachterinnen/Gutachter

1. Prof. Dr. Matthias Jerusalem
2. Prof. Dr. Sigrid Blömeke
3. Prof. Dr. Dieter Kleiber







## **Danksagung**

Während des gesamten Entstehungsprozesses dieser Arbeit habe ich in ganz unterschiedlicher Art und Weise von zahlreichen Personen Unterstützung erhalten.

Allen voran möchte ich meinem Betreuer und Gutachter Prof. Dr. Matthias Jerusalem meinen besonderen Dank aussprechen. Ohne seine Unterstützung und Beratung bei allen fachlichen, konzeptionellen und formalen Fragen sowie ohne seine wertschätzende Teilnahme an allen Höhen und Tiefen im Schreibprozess hätte diese Arbeit nie entstehen können. Auch bei meiner Gutachterin Prof. Dr. Sigrid Blömeke möchte ich mich herzlich bedanken: In ihrem Promotionskolloquium erhielt ich zahlreiche wertvolle Anregungen und Rückmeldungen, die mir bei der Überwindung von Schwierigkeiten geholfen haben. Ihre stetige anspornende Ermutigung war für mich zudem eine wichtige Stütze. Bedanken möchte ich mich ebenfalls bei meinem Gutachter Prof. Dr. Dieter Kleiber: Als studentische Mitarbeiterin habe ich bei ihm bereits die Grundlagen wissenschaftlichen Arbeitens gelernt und konnte auch während der Anfertigung dieser Arbeit immer wieder auf seinen fachlichen Rat zurückgreifen. Ihm und Prof. Dr. Anna Auckenthaler gilt mein besonderer Dank.

Die vorliegende Arbeit ist im Interventionsprojekt „FOSIS2 – Förderung von Schutzfaktoren in der Schule 2“ angesiedelt. FOSIS2 ist ein vom Berliner Senat unterstütztes Kooperationsprojekt der Humboldt-Universität zu Berlin und der Unfallkasse Berlin, das unter der Leitung von Prof. Dr. Matthias Jerusalem an der Abteilung Pädagogische Psychologie und Gesundheitspsychologie durchgeführt wurde. Bei allen Kolleginnen und Kollegen der Abteilung möchte ich mich herzlich bedanken: Ohne den fachlichen Austausch und die Kreativität bei der Bewältigung so mancher organisatorischer Herausforderung hätte ich diese Arbeit nicht verfassen können. Ganz besonders danke ich Merle Kleinfeld für ihr stets offenes Ohr und ihren Humor, Kasra Keshavarz für durchwachte Nächte des gemeinsamen Rechnens und unerwartetes Flow-Erleben beim Eintauchen in das Verfahren der Mehrebenenanalyse sowie Julian Hebenstreit für seine Unterstützung bei der Anfertigung der Tabellen und Grafiken und die wunderbar gelebte Nachbarschaft. Darüber hinaus danke ich Dr. Dietmar Kleine, der mit kriminalistischem Spürsinn Iterationsprobleme gelöst, methodische Fragen beantwortet und die Arbeit Korrektur gelesen hat. Bedanken möchte ich mich auch bei Annette Richter, die immer für mich ansprechbar war und mir bei sämtlichen organisatorischen Fragen ihre warmherzige Hilfe zuteilwerden ließ. Mein Dank gilt des Weiteren allen Projektschülerinnen, -schülern, -lehrerinnen und -lehrern, die mit uner-

müdlichem Einsatz und beispielloser Geduld über zwei Schuljahre hinweg zahlreiche Fragebögen ausgefüllt und die FOSIS2-Interventionen erprobt und umgesetzt haben. Nur durch ihr Engagement war es möglich, das FOSIS2-Projekt durchzuführen und die Fragestellungen dieser Arbeit zu untersuchen.

Eine besondere Erwähnung soll hier mein wissenschaftlicher Mentor, Dr. Dr. Burkhard Gusy finden, dessen wertvolle Rückmeldungen den gesamten Arbeitsprozess begleitet haben und der stets in gedachter Miniaturgröße auf meinem Laptopdeckel sitzt, um mir gedanklich Antworten auf meine Fragen zuzurufen – danke!

Nicht zuletzt möchte ich mich bei all meinen Freunden bedanken, die mir mit ihren Anregungen und Ermutigungen eine unschätzbare Hilfe waren. Speziell danke ich Dr. Dorothea Trebesius und Dr. Guy Tourlamain für ihre Korrekturvorschläge sowie Dr. Rafael Ugarte Chacón für sein großartiges Geschenk der Endkorrektur. Mit Letzterem und auch mit Christina Mayer habe ich die Promotionszeit zu großen Teilen gemeinsam durchlebt, weshalb ihnen mein ganz besonderer Dank gilt. Aufgrund ihres Interesses, ihrer unaufhörlichen Anteilnahme und ihres feinen Gespürs, in den Mühen der Ebene das Skurrile zu entdecken, wird das Abfassen dieser Arbeit eine einzigartige, unvergessliche Zeit bleiben.

Mein größter Dank gilt meiner Familie, speziell meinen Eltern Karola und Horst Chwallek. Ohne ihren bedingungslosen Rückhalt und die unermüdlichen Korrekturvorschläge meines Vaters, der das FOSIS2-Projekt inzwischen wahrscheinlich besser kennt als jeder andere, wäre diese Arbeit nie entstanden. Liebevollster Ratgeber und zugleich schärfster Kritiker war mein Mann Jonas Frister, mit dem ich jede Schwierigkeit diskutieren durfte und in dem ich stets einen geduldigen Gegenleser fand.

Meinen Eltern, Jonas und unserer Tochter Clara, die von Geburt an die Dissertationsschrift als große Schwester akzeptiert hat, ist diese Arbeit gewidmet.

## **Zusammenfassung**

Im Rahmen einer über zwei Schuljahre andauernden quasi-experimentellen Interventionsstudie wurden Lehrkräfte dreier Berliner Oberschulen geschult, durch ein modifiziertes Unterrichtsverhalten und die Einführung von Förderstrategien in ihren Fachunterricht motivational relevante Ressourcen und Risikofaktoren ihrer Schülerinnen und Schüler<sup>1</sup> positiv zu beeinflussen (schulische Selbstwirksamkeitserwartung, Prüfungsängstlichkeit, schulbezogenes Stresserleben und wahrgenommene Binnendifferenzierung im Unterricht). Ziele der vorliegenden Untersuchung sind die Evaluation der Wirksamkeit der Maßnahmen im Hinblick auf die genannten Zielkriterien sowie die Analyse der Effekte in Abhängigkeit von der durch die Schüler wahrgenommenen Umsetzung der Interventionen.

Zu Beginn und am Ende des ersten sowie am Ende des zweiten Projektjahres wurden drei standardisierte Befragungen mit 570 Siebt- und Achtklässlern von drei Interventions- und drei Kontrollschulen durchgeführt. An allen drei Befragungen nahmen 314 Schüler teil (157 Schüler aus den Interventions- und 157 aus den Kontrollschulen).

Die mehrbenenanalytische Untersuchung der Programmeffekte belegt am Ende des ersten Projektjahres bei den Interventionsschülern erwartungskonform günstigere Entwicklungen in der Prüfungsängstlichkeit, der wahrgenommenen schulbezogenen Bedrohung – einer Unterkomponente des Stresserlebens – und der wahrgenommenen Binnendifferenzierung: Bei den Interventionsschülern fallen Prüfungsängstlichkeit und Bedrohungserleben signifikant geringer und die wahrgenommene Binnendifferenzierung signifikant höher aus als bei den Kontrollschülern. Darüber hinaus zeigen sich positive Zusammenhänge zwischen dem Ausmaß der von den Schülern wahrgenommenen Umsetzung der Interventionen und weiteren untersuchten Variablen (z. B. der schulischen Selbstwirksamkeitserwartung).

Die Ergebnisse weisen somit auf die motivationale Relevanz der Fördermaßnahmen speziell des ersten Projektjahres hin. Zudem verdeutlichen die Analysen, dass bei schulischen Interventionen durch Lehrkräfte als Multiplikatoren nicht von einer einheitlichen Umsetzung ausgegangen werden kann und Veränderungen in der Interventionsgruppe auch vom Ausmaß der wahrgenommenen Umsetzung abhängen.

---

<sup>1</sup> Nachfolgend wird zur Verbesserung der Lesbarkeit einheitlich die männliche Form (Schüler, Lehrer) verwendet. Die Begriffe Schüler und Lehrer beinhalten im Weiteren die jeweilige weibliche Form.

## **Abstract**

In the context of a quasi-experimental research project, conducted over two years, teachers of three Berlin secondary schools were trained in new teaching strategies to positively influence students' motivational resources and risk factors. The aim of this study is to evaluate the program's effects on students' academic self-efficacy, test anxiety, school-related stress as well as internal differentiation and to analyze these effects as a function of students' perception of the program's implementation.

At the beginning of the first project year and at the end of the first and second project years standardized surveys were conducted in three experimental schools and three control schools with 570 students in the seventh and eighth grades. A total of 314 students took part in all three surveys (157 students in the experimental group and 157 students in the control group).

As expected, multilevel analysis showed significant effects in students' test anxiety, perceived threat – a component of stress – as well as internal differentiation at the end of the first project year: Test anxiety and perceived threat were significantly lower in the experimental group than in the control group, whereas internal differentiation was higher in the experimental group. In addition correlations were found between student's perception of the program's implementation and several other variables under study (e.g. self-efficacy).

The results therefore underline the program's relevance for motivational resources and risk factors. This is particularly true for the first year interventions. Furthermore, the results indicate that program implementation varies between teachers and that implementation quantity considerably affects program effects.



# Inhaltsverzeichnis

<b>Danksagung.....</b>	<b>i</b>
<b>Zusammenfassung.....</b>	<b>iii</b>
<b>Abstract.....</b>	<b>iv</b>
<b>1 Einleitung.....</b>	<b>1</b>
<b>2 Motivation in der Schule .....</b>	<b>7</b>
<b>2.1 Definition von Lern- und Leistungsmotivation.....</b>	<b>8</b>
<b>2.2 Schulische Selbstwirksamkeitserwartung – eine Ressource für die         Motivation .....</b>	<b>10</b>
2.2.1 Definition und theoretische Konzeption von Selbstwirksamkeitserwartung....	11
2.2.2 Entstehungsbedingungen von Selbstwirksamkeitserwartung.....	12
2.2.3 Auswirkungen und Korrelate von Selbstwirksamkeitserwartung .....	14
2.2.4 Mit Selbstwirksamkeitserwartung verwandte Konstrukte.....	17
2.2.5 Erfassung von Selbstwirksamkeitserwartung.....	20
2.2.6 Implikationen für die Praxis: Fördermöglichkeiten von Selbstwirksamkeitserwartung.....	20
2.2.7 Zusammenfassung: Selbstwirksamkeitserwartung.....	24
<b>2.3 Prüfungsängstlichkeit – ein Risikofaktor für die Motivation.....</b>	<b>25</b>
2.3.1 Definition und theoretische Konzeption von Prüfungsangst und -ängstlichkeit.....	26
2.3.2 Entstehungsbedingungen von Prüfungsangst und -ängstlichkeit .....	27
2.3.3 Auswirkungen und Korrelate von Prüfungsangst- und -ängstlichkeit.....	29
2.3.4 Mit Prüfungsangst und -ängstlichkeit verwandte Konstrukte .....	32
2.3.5 Erfassung von Prüfungsangst und -ängstlichkeit.....	33
2.3.6 Implikationen für die Praxis: Reduktion von Prüfungsangst und -ängstlichkeit.....	34
2.3.7 Zusammenfassung: Prüfungsangst und -ängstlichkeit.....	39

<b>2.4 Schulbezogenes Stresserleben – Herausforderungserleben als Ressource, Bedrohungs- und Verlusterleben als Risikofaktoren für die Motivation.....</b>	<b>40</b>
2.4.1 Definition und theoretische Konzeption von Stresserleben .....	42
2.4.2 Entstehungsbedingungen von Stresserleben .....	47
2.4.3 Auswirkungen und Korrelate von Stresserleben .....	49
2.4.4 Ressourcen und Stressbewältigung .....	52
2.4.5 Angst und Stresserleben als verwandte Konstrukte .....	54
2.4.6 Erfassung von Stresserleben.....	55
2.4.7 Implikationen für die Praxis: Reduktion von Stresserleben.....	55
2.4.8 Zusammenfassung: Stresserleben.....	59
<b>2.5 Strategien und Ansätze zur Motivationsförderung in der Schule .....</b>	<b>60</b>
2.5.1 Binnendifferenzierung.....	61
2.5.2 Interventionsstudien zur Motivationsförderung in der Schule .....	63
2.5.3 Zusammenfassung: Strategien und Ansätze zur Motivationsförderung in der Schule .....	67
<b>3 Evaluations- und Implementationsforschung .....</b>	<b>68</b>
<b>3.1 Evaluationsforschung.....</b>	<b>69</b>
3.1.1 Definition von Evaluation .....	69
3.1.2 Funktionen von Evaluationen.....	70
3.1.3 Ziele von Evaluationen.....	71
3.1.4 Rahmenbedingungen und Merkmale von Evaluationen .....	72
3.1.5 Standards von Evaluationen .....	74
3.1.6 Wirksamkeitsevaluation .....	75
3.1.7 Zusammenfassung: Evaluationsforschung.....	79
<b>3.2 Implementationsforschung.....</b>	<b>80</b>
3.2.1 Definition von Implementation .....	81
3.2.2 Implementationskennwerte .....	82
3.2.3 Implementationsstrategien.....	83

3.2.4	Methoden zur Erfassung der Implementation.....	85
3.2.5	Rahmenbedingungen und Merkmale für eine erfolgreiche Implementation....	85
3.2.6	Zusammenfassung: Implementationsforschung .....	88
<b>4</b>	<b>Das FOSIS2-Projekt.....</b>	<b>90</b>
4.1	Bezug und Abgrenzung zu Vorgängerprojekten .....	90
4.2	Zeitlicher Verlauf des Projekts.....	94
4.3	Beschreibung der Module .....	96
4.3.1	Klassenklima.....	97
4.3.2	Transparenz.....	100
4.3.3	Lern- und Leistungsraum.....	102
4.3.4	Selbstbestimmung.....	104
4.3.5	Kooperatives Lernen.....	105
4.3.6	Selbstregulation .....	107
4.4	Zusammenfassung der Projektbeschreibung.....	110
<b>5</b>	<b>Fragestellungen .....</b>	<b>111</b>
5.1	Wirksamkeitsevaluation.....	111
5.1.1	Trägt die Intervention zu einer günstigeren Entwicklung der motivationalen Ressourcen und Risikofaktoren bei? .....	112
5.1.2	Gibt es jahresspezifische Effekte? .....	114
5.2	Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation .....	115
5.2.1	Hängt die Wirksamkeit der Intervention mit dem von den Schülern wahrgenommenen Ausmaß der Implementation zusammen? .....	115
<b>6</b>	<b>Methode.....</b>	<b>116</b>
6.1	Durchführung des FOSIS2-Projekts.....	116
6.1.1	Auswahl und Rekrutierung der Stichprobe.....	116
6.1.2	Durchführung der Workshops, Beratungen und Befragungen .....	118
6.1.3	Besondere Begleitumstände.....	121

<b>6.2 Design.....</b>	<b>123</b>
6.2.1 Design zur Evaluation der Wirksamkeit .....	123
6.2.2 Design zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation.....	124
<b>6.3 Stichprobe .....</b>	<b>125</b>
6.3.1 Beschreibung der Gesamtstichprobe .....	125
6.3.2 Umgang mit fehlenden Werten und Wave-Nonresponse-Analysen .....	128
6.3.3 Beschreibung der Analytestichproben .....	136
<b>6.4 Erhebungsinstrumente.....</b>	<b>138</b>
6.4.1 Skalen der Prä-, Zwischen- und Postbefragung .....	138
6.4.2 Skalen der modulspezifischen Rückmeldungen.....	141
<b>6.5 Methodisches Vorgehen: Mehrebenenanalysen .....</b>	<b>143</b>
6.5.1 Gründe für die Anwendung von Mehrebenenanalysen .....	144
6.5.2 Veränderungsmessung im Rahmen von Mehrebenenanalysen .....	146
6.5.3 Mehrebenenmodelle zur Evaluation der Wirksamkeit.....	148
6.5.4 Mehrebenenmodelle zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation .....	158
6.5.5 Zusammenfassung: Methodisches Vorgehen – Mehrebenenanalysen .....	164
<b>7 Ergebnisse .....</b>	<b>165</b>
<b>7.1 Wirksamkeitsevaluation .....</b>	<b>165</b>
7.1.1 Lineare Wachstumskurvenmodelle .....	166
7.1.2 Diskontinuierliche Veränderungsmodelle .....	178
7.1.3 Zusammenfassung der Ergebnisse der Wirksamkeitsevaluation .....	194
<b>7.2 Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation.....</b>	<b>195</b>
7.2.1 Deskriptive Beschreibung der Implementation .....	195
7.2.2 Kovarianzanalytische Veränderungsmodelle .....	198

7.2.3 Zusammenfassung der Ergebnisse der Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation .....	217
<b>8 Diskussion .....</b>	<b>218</b>
8.1 Diskussion der Ergebnisse zur Wirksamkeitsevaluation .....	218
8.2 Diskussion der Ergebnisse zur Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation.....	226
8.2.1 Deskriptive Beschreibung der Implementation .....	226
8.2.2 Kovarianzanalytische Veränderungsmodelle .....	230
8.3 Bemerkungen zur Durchführung des FOSIS2-Projekts.....	232
8.4 Einschränkungen der Studie .....	234
8.5 Zusammenfassung und Ausblick.....	238
<b>Literaturverzeichnis.....</b>	<b>242</b>
<b>Tabellenverzeichnis.....</b>	<b>263</b>
<b>Abbildungsverzeichnis.....</b>	<b>265</b>
<b>Anhang .....</b>	<b>267</b>



# 1 Einleitung

*„Die Schule ist ein Gefängnis für alle, die ihre Möglichkeiten nicht für sich nutzen können. Andere wieder können selbst in der gefängnisähnlichsten Schule Sinn und Zweck für ihre Arbeit erkennen. Um seine Schulzeit als sinnvoll erleben zu können, braucht das Kind konkrete persönliche Gründe für seinen Schulbesuch überhaupt, ein Bedürfnis nach dem, was die Schule zu bieten hat, kurz: Lernmotivation. Ist das Kind mit persönlichem Engagement bei der schulischen Arbeit, kommt es auch aus eigenem Antrieb zur Schule, eben weil es den Unterricht als sinnvoll erlebt“ (DeCharms, 2011, S. 25).*

Die Motivation ist ein umgangssprachlich und auch medial vielfach thematisierter Begriff (Heckhausen, 2010), ein schillerndes, sozial erwünschtes Konzept, dem eine Schlüsselfunktion für den Erfolg in schulischen und beruflichen Kontexten zugeschrieben wird. So veröffentlichte beispielsweise *Der Spiegel* im Januar 2013 einen Artikel mit dem Titel „Erfolg in Mathe: Motivation ist wichtiger als Intelligenz“ (Dambeck, 2013). In der *ZEIT* erschien, ebenfalls im Januar 2013, ein Beitrag mit dem Titel „Was Mitarbeiter wirklich motiviert“ (Kauß, 2013) und die *Frankfurter Allgemeine* empfahl im Februar desselben Jahres in einem Bericht zum erfolgreichen Lernen nicht nur Lerntechniken, sondern auch Motivationstipps (Bös, 2013). Auch zahlreiche Ratgeber (z. B. Brohm, 2011; Endres, 2012; Fräd- rich, 2006; Schmidt, 2010) und Internetseiten (z. B. [www.motivationstipps.de](http://www.motivationstipps.de), <http://motivationstraining.wordpress.com>; 24.01.2015) bieten ein breites Spektrum an Tipps zur Selbstmotivierung und zur Motivierung anderer an.

Im schulischen Kontext stellt die Motivation eine zentrale Ressource für die gleichzeitige und kontinuierliche Bewältigung der zahlreichen, teilweise sehr komplexen schulischen Anforderungen dar: Schüler sollen lernen und Leistungen erbringen, die bestimmte, ihrem Alter entsprechende Kompetenzstandards einhalten. Darüber hinaus wird von den Schülern erwartet, dass sie sich in die Schul- und Klassengemeinschaft integrieren, soziale Beziehungen zu Mitschülern und Lehrkräften knüpfen und aufrechterhalten, sich an Schul- und Klassenregeln halten und an gemeinsamen schulischen Aktivitäten beteiligen. Für die erfolgreiche Bewältigung dieser Anforderungen ist eine hohe Motivation unerlässlich (Wentzel & Wigfield, 2009). Denn diese wirkt sich nicht nur positiv auf Lernen und Leistung aus, sondern trägt auch dazu bei, das schulische Engagement insgesamt zu stärken und die Bewältigungsfähigkeiten sowie die Schüler-Schüler- und Schüler-Lehrer-Beziehungen zu verbessern (Martin, 2008; Schiefele, 2009).

Lehrer aller Schulformen sehen sich jedoch immer wieder mit dem Problem konfrontiert, dass manche Schüler einer Klasse ein hochmotiviertes, engagiertes und involviertes Lernverhalten an den Tag legen und andere, im Gegensatz dazu, unmotiviert und apathisch erscheinen. Diese Differenz im Verhalten von Schülern einer Klasse kann zu einer Verschlechterung des Klassenklimas und der Lernbedingungen durch beispielsweise Unterrichtsstörungen einzelner demotivierter Schüler führen (Linnenbrink & Pintrich, 2003; Schiefele, 2009). Insbesondere nach dem Übergang von der Grund- zur weiterführenden Schule ist über alle Schultypen hinweg eine deutliche Abnahme der Motivation festzustellen (Wigfield, Ho & Mason-Singh, 2011). Die mangelnde Motivation ist somit speziell ein Problem der weiterführenden Schulen, für das bisher keine einfache Lösung gefunden werden konnte (Linnenbrink & Pintrich, 2003; Wigfield, Cambria & Eccles, 2012). Zumindest bietet die anfangs geschilderte Ratgeber-Literatur nur wenig theoretisch fundierte Hilfen, denn oft handelt es sich hierbei um rein erfahrungsbasiertes Wissen, das nicht evaluiert wurde. Auch der Blick in Wissenschaft und Forschung zum Thema Motivation lässt keine einfachen Schlüsse auf den praktischen Umgang mit Motivationsproblemen zu. In vielen Studien der Motivationsforschung wurde zwar auf die Bedeutsamkeit einer Reihe von Variablen hingewiesen, die für die Entwicklung und Aufrechterhaltung der Motivation in der Schule eine große Rolle spielen (z. B. Linnenbrink & Pintrich, 2003; Martin, 2008; Wigfield et al., 2011). Die psychologische Motivationsforschung ist jedoch sehr breit gefächert und betrifft zahlreiche verschiedene Konzepte (Ryan, 2012), sodass nach Martin (2008) im Hinblick auf die Frage nach praxiswirksamen Maßnahmen zur Motivationsförderung eine Verunsicherung entsteht und die bisherige Forschung nur wenige wirklich anwendbare Implikationen für die Schulpraxis zulässt. Genau dies stellt jedoch das Anliegen der Lehrkräfte dar, die angesichts der Motivationsprobleme ihrer Schüler daran interessiert sind, Wege und Strategien zu finden, wie sie die Motivation ihrer Schüler so erzeugen beziehungsweise verstärken können (Fries & Souvignier, 2009; Rheinberg & Krug, 2005b), dass die Schüler irgendwann in der Lage sind, sich selbst zu motivieren (Helmke, 2014). Immer wieder wird zur Verringerung der vorherrschenden Verunsicherung über praktische Maßnahmen zur Motivationsförderung im Bereich Schule und Lernen gefordert, künftig mehr Längsschnittuntersuchungen durchzuführen (Ryan & Legate, 2012) und eine stärker interdisziplinäre Herangehensweise sowie ein multidimensional ausgerichtetes Vorgehen zu wählen, das unterschiedliche theoretische Perspektiven und verschiedene motivational relevante Konzepte vereint. Psychologische und pädagogische Erkenntnisse sollen miteinander verbunden und das Zusammenspiel zentraler Einflussfaktoren auf die



Motivation geprüft werden (Martin, 2008; Pintrich, 2003). Angesichts der Vielzahl lernrelevanter Motivationsgrößen ist es nicht möglich, das Zusammenwirken aller für die Motivation einer Person maßgeblichen Komponenten gleichzeitig zu berücksichtigen (Wild, Hofer & Pekrun, 2006). Zumindest aber sollen Schutzfaktoren beziehungsweise Ressourcen, die für den Aufbau und die Aufrechterhaltung von Motivation eine zentrale Rolle spielen, genauso berücksichtigt werden wie Risikofaktoren, die den Aufbau einer Motivation verhindern oder eine bereits vorhandene Motivation verringern. Auch werden mehr Studien gewünscht, die gleichzeitig intrapersonale psychische Prozesse und situative beziehungsweise kontextspezifische Einflüsse auf die Motivation untersuchen (Martin, 2008; Pintrich, 2003). Zudem wird der Mangel an theorie-basierter Anwendungsforschung beklagt: Förderprogramme sollen auf der Basis der Erkenntnisse der motivationspsychologischen Grundlagenforschung konzipiert, in der Schule angewendet und evaluiert werden (2008; Pintrich, 2003; Rheinberg & Krug, 2005b). Auch Wigfield, Cambria und Eccles (2012) ermutigen im „The Oxford Handbook of Human Motivation“ zu einer intensiveren Interventionsforschung im Rahmen quasi-experimenteller Studien in der schulischen Praxis. Obwohl bei der Umsetzung solcher Interventionsstudien im Kontext Schule mit Schwierigkeiten und Hindernissen zu rechnen ist, sind solche Interventionsstudien unbedingt erforderlich (Linnenbrink & Pintrich, 2002; Rheinberg & Krug, 2005b). Denn verschiedene Förderstrategien können sich in Abhängigkeit von Merkmalen eines Schülers und Besonderheiten seiner Klasse und Schule auch ganz unterschiedlich auswirken (Wigfield et al., 2012). Erst die Evaluation von Interventionen in realen Schulen unter Berücksichtigung der spezifischen Kontextbedingungen kann wirkliche Erkenntnisse über die Wirksamkeit von Motivationsfördermaßnahmen in der Praxis liefern und helfen, die Lücke zwischen pädagogisch-psychologisch theoretischen Konzepten auf der einen Seite und den Lehrkräften und deren Anliegen auf der anderen Seite zu schließen (Linnenbrink & Pintrich, 2002; Rheinberg & Krug, 2005b). In methodischer Hinsicht ist bei der Evaluation solcher Förderprogramme im Rahmen quasi-experimenteller Studien unbedingt zu beachten, dass die intendierten Wirkungen insbesondere bei komplexen, aus mehreren Modulen bestehenden Interventionen eng mit der Programmumsetzung, der Implementation, zusammenhängen (Durlak & DuPre, 2008; Kalafat, Illback & Sanders, 2007). Für die Programmwirksamkeit ist entscheidend, wie und unter welchen Bedingungen die einzelnen Programmkomponenten in die Praxis implementiert wurden (Gollwitzer & Jäger, 2009), denn eine unzureichende Implementation kann die Programmwirkung schwächen. Insbesondere wenn Interventionen durch vermittelnde Personen (Mediatoren), wie z. B. Lehr-

kräfte, umgesetzt werden, ist eine Evaluation der Implementation unerlässlich, da Ausmaß und Genauigkeit der Umsetzung zwischen den Mediatoren variieren können (Durlak, 2013; Durlak & DuPre, 2008; Petermann, 2014b). Obwohl zahlreiche Studien belegen, dass bei der Implementation von Interventionen in der Schule mit besonderen Schwierigkeiten zu rechnen ist (z. B. Zeitprobleme, mangelnde Unterstützung durch die Schulleitung), erfolgt eine Evaluation der Umsetzung eher selten (Payne, 2009; Payne, Gottfredson & Gottfredson, 2006; Resnicow et al., 1992; Scheirer, Shediak & Cassady, 1995). Insbesondere bei schulbasierten Interventionen ergibt sich daher die Notwendigkeit, die Wirksamkeit auch in Abhängigkeit von der Implementation zu evaluieren (Durlak & DuPre, 2008; Fries & Souvignier, 2009; Gollwitzer & Jäger, 2009; Gräsel & Parchmann, 2004; Hasselhorn, Köller, Maaz & Zimmer, 2014).

Zur Klassifikation und Einordnung von Motivationsförderprogrammen für die Schule treffen Rheinberg und Krug (2005b) folgende Unterscheidungen: Ein Programm kann an der Person ansetzen und versuchen, intrapersonale Prozesse zur Stärkung der Motivation zu verändern; oder eine Intervention zielt darauf ab, die Schüler durch situative Veränderungen zu motivieren.<sup>2</sup> Im besten Fall berücksichtigt ein Programm, wie oben gefordert, gleichermaßen intrapersonale und situative Einflussfaktoren. Zudem kann ein Programm zum Ziel haben, kurzfristig auf die Motivation Einfluss zu nehmen, beispielsweise zu Beginn der Stunde zu motivieren. Ein Motivationsförderprogramm kann aber auch beabsichtigen, eine Motivation längerfristig aufrechtzuerhalten (z. B. einen insgesamt wenig lernmotivierten Schüler in seiner Lernmotivation zu bestärken). Schließlich halten Rheinberg und Krug (2005b) noch fest, dass sich die Maßnahmen zur Förderung der Motivation in Abhängigkeit von ihrer psychologischen Grundposition auch in ihren Wirkmechanismen unterscheiden. So erklären behavioristisch angelegte Förderprogramme Verhalten als Reaktion auf bestimmte Bedingungen, die es zu verändern gilt, um die Motivation zu verbessern. Motivationstheoretisch verankerte Förderprogramme hingegen „betonen stärker die Eigeninitiative der Person“ (Rheinberg, 2005, S. 18) und erklären Verhalten und Handeln durch zukünftig erwartete Zustände oder Ergebnisse, die positive Folgen nach sich ziehen. Solche Förderprogramme zielen auf die Veränderung von Erwartungen und Folgeannahmen ab, um die Motivation zu verstärken.

---

<sup>2</sup> Diese Unterscheidung entspricht der Unterteilung von präventiven Interventionen in Verhaltens- und Verhältnispräventionen. Erstere Präventionsform zielt auf die Verhaltensveränderung eines Individuums ab. Bei letzterer sollen Umwelt- und Lebensbedingungen verändert werden (Kleiber, 2012).

Die vorliegende Arbeit ist im Interventionsprojekt „FOSIS2 – Förderung von Schutzfaktoren in der Schule 2“ angesiedelt.<sup>3</sup> FOSIS2 ist ein vom Berliner Senat unterstütztes Kooperationsprojekt der Humboldt-Universität zu Berlin und der Unfallkasse Berlin, das unter der Leitung von Prof. Dr. Matthias Jerusalem an der Abteilung Pädagogische Psychologie und Gesundheitspsychologie durchgeführt wurde. Im Rahmen dieser über zwei Schuljahre andauernden quasi-experimentellen Längsschnittstudie wurden Lehrkräfte als Mediatoren geschult, durch ein modifiziertes Unterrichtsverhalten und die Einführung von Förderstrategien in ihren regulären Fachunterricht die motivationalen und sozialen Ressourcen und Risikofaktoren ihrer Schüler positiv zu beeinflussen. Zielgruppe waren Schülerinnen und Schüler der siebten und achten Jahrgänge dreier Berliner Oberschulen. Drei weitere Oberschulen nahmen als Kontrollschulen am Projekt teil. In der vorliegenden Untersuchung liegt der Fokus auf den motivationalen Effekten der FOSIS2-Fördermaßnahmen. Im Hinblick auf die von Rheinberg und Krug (2005b) beschriebenen Unterscheidungen zur Einordnung von Interventionen zur Motivationsförderung lässt sich das FOSIS2-Programm folgendermaßen charakterisieren: Das Programm enthält sowohl Maßnahmen zur Veränderung personaler Merkmale der Schüler (z. B. Aufbau von Kompetenzüberzeugungen und Reduktion von Prüfungsängstlichkeit) als auch Maßnahmen zur Veränderung des Lehrerverhaltens und der schulischen Kontextbedingungen (z. B. Steigerung der Binnendifferenzierung). Zudem hat das Programm zum Ziel, eher langfristig die Lernmotivation der Schüler zu stärken und ist motivationstheoretisch verankert. Die Konzeption der FOSIS2-Maßnahmen basiert jedoch nicht allein auf psychologischen Erkenntnissen, sondern auch auf erziehungswissenschaftlichen. Wie gefordert werden folglich interdisziplinäre Perspektiven berücksichtigt. Die Intervention ist multidimensional angelegt und soll bei den Interventionsschülern verglichen mit der Kontrollgruppe zu einer günstigeren Entwicklung von folgenden Ressourcen und Risikofaktoren für die Motivation beitragen: Selbstwirksamkeitserwartung, Prüfungsängstlichkeit, Stresserleben und wahrgenommene Binnendifferenzierung im Unterricht. Das FOSIS2-Programm wird im Schulalltag erprobt und im Rahmen einer quasi-experimentellen Längsschnittstudie evaluiert. Somit greift das Programm die oben geschilderten Forderungen an die Motivationsforschung auf und stellt einen Versuch dar, neue Erkenntnisse über theoretisch-fundierte, für die schulische Praxis

---

<sup>3</sup> FOSIS2 ist ein Nachfolgeprojekt von FOSIS1, das in den Jahren 2006 und 2007 von den gleichen Kooperationspartnern durchgeführt wurde. Die Unterschiede der beiden Projekte werden in Abschnitt 4.1 beschrieben.

anwendbare Strategien zur Motivationsförderung zu ermitteln. Inwieweit dies tatsächlich gelingt, soll in der vorliegenden Arbeit beantwortet werden.

Die Arbeit beginnt mit der Darstellung des motivationstheoretischen Hintergrunds (Kapitel 2): Zunächst werden die für die Schule bedeutsamen Motivationsformen, die Lern- und die Leistungsmotivation, definiert. Anschließend werden die relevanten Zielkriterien des FOSIS2-Programms, die schulische Selbstwirksamkeitserwartung als Ressource für die Motivation und die Prüfungsängstlichkeit sowie das negative schulbezogene Stresserleben als Risikofaktoren für die Motivation, geschildert. Das Kapitel schließt mit einer Auswahl an Strategien und Ansätzen zur Förderung der Motivation. In dem Zusammenhang wird auch ein weiteres Zielkriterium des FOSIS2-Programms beschrieben: die Binnendifferenzierung im Unterricht als eine motivational relevante, externe Ressource der Lernumwelt. Im dritten Kapitel werden Erkenntnisse aus der Evaluations- und Implementationsforschung zusammengefasst, und im Anschluss widmet sich das vierte Kapitel dem FOSIS2-Projekt. Zunächst folgt eine Abgrenzung zu Vorgängerprojekten. Daran anknüpfend werden der zeitliche Ablauf des FOSIS2-Projekts und die Inhalte der Intervention erläutert. Nach der Projektbeschreibung werden im fünften Kapitel die Fragestellungen und die Forschungshypothesen vorgestellt. Zum einen wird der Frage nach der Wirksamkeit der FOSIS2-Maßnahmen im Hinblick auf die genannten motivational relevanten Variablen nachgegangen. Zum anderen soll analysiert werden, inwiefern die Wirksamkeit mit der von den Schülern wahrgenommene Umsetzung der FOSIS2-Maßnahmen in den Unterricht zusammenhängt. Unter dem Titel „Methode“ wird in Kapitel 6 zunächst die Durchführung des Projekts, also die Auswahl und Rekrutierung der Stichprobe sowie die Durchführung der Workshops, Beratungen und Befragungen beschrieben. Auch das quasi-experimentelle Studien-Design, Merkmale der Untersuchungsstichproben und das methodische Vorgehen werden im sechsten Kapitel erläutert. Zur Beantwortung der Fragestellungen wurde das Verfahren der Mehrebenenanalyse herangezogen. Dabei werden neben üblichen Formen der Mehrebenenmodellierung, wie beispielsweise den linearen Wachstumskurvenmodellen, auch sogenannte diskontinuierliche Modelle eingesetzt. Da diese im Rahmen manifester Mehrebenenanalysen eine nur selten gewählte Form der Modellierung darstellen (Luhmann & Eid, 2013), wird das methodische Vorgehen entsprechend ausführlich erläutert. Im siebten Kapitel werden die Ergebnisse der Wirksamkeitsevaluation und der Implementationsanalysen erläutert, die dann schließlich im achten und letzten Kapitel diskutiert und interpretiert werden. Die Arbeit schließt mit einer Zusammenfassung der Ergebnisse und einem Ausblick.

## 2 Motivation in der Schule

Der Begriff der Motivation erweckt durch seinen häufigen Gebrauch in Umgangssprache und Alltag den Eindruck, eine klare, eindeutige und allgemeingültige Bedeutung aufzuweisen. Tatsächlich scheint die eigene Motivation auch unmittelbar erfahrbar zu sein (DeCharms, 2011) und kann in ihrer Intensität (hoch, niedrig) variieren (Rheinberg & Vollmeyer, 2012). Die Motivation anderer ist jedoch nicht direkt beobachtbar, sondern muss aus dem Verhalten des anderen erschlossen werden. Verhalten lässt sich allerdings mit Motivation nicht immer gleichsetzen, sodass Rückschlüsse vom Verhalten auf die Motivation des anderen falsch sein können (DeCharms, 2011). „So aussagekräftig der Begriff Motivation für uns subjektiv sein mag – als wissenschaftlicher Terminus ist er unzureichend definiert“ (DeCharms, 2011, S. 18). Motivation ist folglich kein eindeutig definierter Zustand, sondern vielmehr ein abstrakter Oberbegriff, der zahlreiche verschiedene Komponenten und Prozesse unseres menschlichen Handelns vereint, die gemeinsam versuchen zu begründen, warum ein bestimmtes Verhalten aufgenommen und aufrechterhalten und ein anderes unterlassen wird (Rheinberg & Vollmeyer, 2012; Ryan, 2012). Mit dem Zustand des „Motiviert-Seins“ werden umgangssprachlich Prozesse wie Streben, Wollen, Wünschen, Verlangen und damit verbundene Verhaltensauswirkungen wie die Zielverfolgung, Anstrengung und Ausdauer assoziiert (Rheinberg & Vollmeyer, 2012). Aus wissenschaftlicher Perspektive umfasst der Begriff der Motivation weitere zahlreiche Komponenten wie beispielsweise Einstellungen, Überzeugungen, Erwartungen, Wertvorstellungen, Attributionen und Emotionen (Wentzel & Wigfield, 2009). Nach Rheinberg und Vollmeyer (2012) beschreiben alle diese Konzepte gemeinsam „die aktivierende Ausrichtung des Lebensvollzugs auf einen positiv bewerteten Zielzustand“ (S. 15). Die Motivation und die sich hinter diesem Begriff verbergenden Prozesse bestimmen folglich die Richtung, die Intensität und die Ausdauer unseres Verhaltens (Rheinberg & Vollmeyer, 2012). Somit ist die Motivation ein dynamisches, facettenreiches Konzept und kein stabiles, einheitliches Merkmal einer Person. Vielmehr ist sie in der Regel innerhalb einer Person situations-, kontext- und domänenspezifisch und kann in ihrer Intensität variieren. Bezogen auf die Schule bedeutet dies, dass die für diesen Kontext relevante Lern- und Leistungsmotivation nicht nur von der Person selbst, sondern auch von den Fächern und der Beschaffenheit der Situation und der Klassenumgebung abhängt. Die Motivation in der Schule zu erfassen und zu beeinflussen, ist folglich ein hohes und komplexes Ziel (Linnenbrink & Pintrich, 2002).

Aufgrund der großen Anzahl der für die Motivation relevanten Konzepte und Prozesse kann im Rahmen empirischer Interventionsstudien zur Motivationsförderung in der Schule immer nur eine Auswahl von Konstrukten untersucht werden. Nach einer Begriffsdefinition von Lern- und Leistungsmotivation werden deshalb hier solche für die Motivation bedeutsamen Konzepte dargestellt, die für die Entwicklung des FOSIS2-Förderprogramms entscheidend waren und die bei der Evaluation der Intervention als Zielkriterien herangezogen werden. Bei der Konzeption des FOSIS2-Projekts wurde die Forderung der Motivationspsychologie aufgegriffen, nicht nur Ressourcen, sondern auch Risikofaktoren für die schulische Motivation zu berücksichtigen. Dieser Forderung entsprechend werden in der vorliegenden Untersuchung die schulische Selbstwirksamkeitserwartung als eine zentrale Ressource für die Motivation und die Prüfungsängstlichkeit sowie das schulbezogene Stresserleben als Risikofaktoren für die Motivation untersucht. Diese drei Zielkriterien werden nachfolgend erläutert. Im Anschluss folgt eine Beschreibung von Strategien und Ansätzen zur Förderung der Motivation, wobei zunächst eine spezifische Unterrichtsstrategie berichtet wird, die ebenfalls durch das FOSIS2-Programm positiv beeinflusst werden soll: die Binnendifferenzierung im Unterricht. Das FOSIS2-Programm soll nämlich nicht nur zu einer Veränderung der benannten motivational relevanten personalen Merkmale der Schüler beitragen, sondern auch das Lehrerverhalten und somit die schulischen Kontextbedingungen beeinflussen, in motivationaler Hinsicht z. B. die Binnendifferenzierung im Unterricht als eine externe Ressource der Lernumwelt. Dementsprechend wird in dieser Arbeit die von den Schülern wahrgenommene Binnendifferenzierung als ein weiteres Zielkriterium untersucht. Das Kapitel 2 schließt mit einer Darstellung ausgewählter Motivationsförderprogramme, die ähnlich wie das FOSIS2-Programm verschiedene Förderstrategien kombinieren und auf die gleichzeitige Verbesserung mehrerer motivational relevanter Variablen abzielen.

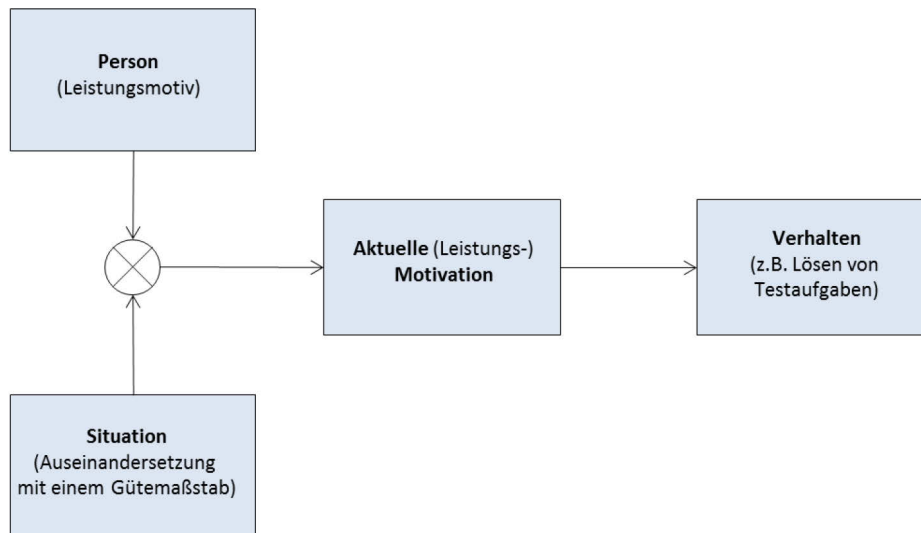
## **2.1 Definition von Lern- und Leistungsmotivation**

Unter der *Lernmotivation* wird der Wunsch oder die Absicht verstanden, eine Lernaktivität aufzunehmen, um einen Lern- oder Kompetenzzuwachs zu erleben (Rheinberg, 1986). Die *Leistungsmotivation*, die deutlich besser erforscht ist, wird umgangssprachlich gerne mit Begriffen wie „Fleiß, Eifer, Arbeitswillen oder auch Strebsamkeit“ (Rheinberg & Vollmeyer, 2012, S. 60) assoziiert. Die durch diese Begriffe erfassten Prozesse sind jedoch nicht zwangsläufig leistungsmotiviert. Vielmehr ist ein Verhalten im wissenschaftlichen Verständnis nur dann leistungsmotiviert, „wenn es in Auseinandersetzung mit einem Gü-

temaßstab, den es zu erreichen oder zu übertreffen gilt, auf die Selbstbewertung der eigenen Tüchtigkeit abzielt“ (Rheinberg & Vollmeyer, 2012, S. 60). Zu den frühen Pionieren und Begründern der Forschung zur Leistungsmotivation zählen vor allem McClelland und Atkinson (z. B. McClelland, Atkinson, Clark & Lowell, 1953; Atkinson, 1957, 1958). In Anlehnung an Lewin (1951), der forderte, Verhalten grundsätzlich als eine Funktion von Person und Umwelt zu betrachten, berücksichtigten McClelland und Atkinson für die Erklärung des Zustandekommens einer aktuellen Leistungsmotivation ebenfalls personale und situative Komponenten. Der Zusammenhang wird in dem Modell in Abbildung 1 veranschaulicht (Rheinberg & Vollmeyer, 2012), das heute als das Grundmodell der klassischen Motivationspsychologie gilt. Hiernach ist für eine aktuelle Leistungsmotivation auf personaler Seite das *Leistungsmotiv* erforderlich. Motive werden aufgefasst als stabile, personale Merkmale, die bestimmen, wie Personen Situationen wahrnehmen und diese im Hinblick auf ihr eigenes Handeln interpretieren. Wenn auf situativer Seite motiv-passende, im Fall der Leistungsmotivation also leistungsthematische *Anreize* hinzukommen, entsteht eine aktuelle Leistungsmotivation. Ermöglicht eine Situation beispielsweise eine Auseinandersetzung mit einem Gütemaßstab, kann sie das Leistungsmotiv einer Person aktivieren und eine aktuelle Leistungsmotivation nach sich ziehen.<sup>4</sup> Dieser Prozess der Aktivierung eines Motivs durch die thematische Passung zwischen Situation und Motiv wird auch *Motivierung* genannt. Personen können sich hinsichtlich der Ausprägung ihrer Motive und somit auch in ihrer Interpretation von Situationen unterscheiden. So nehmen Personen mit einem hohen Leistungsmotiv eher Gelegenheiten wahr, bei denen sie ihre Fähigkeiten erproben und demonstrieren können. Im Gegensatz dazu bemerken Personen mit einem stark ausgeprägten Machtmotiv in Situationen eher, wann sich Möglichkeiten der Einflussnahme und des Machtgewinns ergeben. Für Personen mit einem hohen Anschlussmotiv besitzen vor allem soziale Situationen einen Aufforderungscharakter. Ein und dieselbe Situation kann durch die Motivausprägung somit ganz unterschiedlich interpretiert werden (Rheinberg & Vollmeyer, 2012).

---

<sup>4</sup> Scheffer und Heckhausen (2010) beschreiben fünf Kriterien, die erforderlich sind, um eine Situation als leistungsthematisch bewerten zu können: Die Handlung in der Situation sollte mit einem nachweisbaren Ergebnis beendet werden (1). Dieses Ergebnis sollte mittels vor der Handlung definierter Gütemaßstäbe oder im Hinblick auf seine Quantität bewertbar sein (2). Hinsichtlich der Schwierigkeit sollte die Handlung Anforderungen beinhalten, die weder zu leicht noch zu schwierig sind (3). Zudem sollten Vergleichsmaßstäbe existieren, die eine Bewertung der Handlung erleichtern (4). Schließlich muss die Handlung vom Handelnden gewünscht und selbst ausgeführt sein (5).



**Abbildung 1: Grundmodell der klassischen Motivationspsychologie nach Rheinberg & Vollmeyer (2012), S. 70**

Das klassische Grundmodell der Motivationspsychologie ist in den letzten fünfzig Jahren zunehmend differenziert und im Zuge der kognitiven Wende vor allem durch kognitive Prozesse wie Erwartungen und Überzeugungen erweitert worden (Rheinberg & Vollmeyer, 2012). Eine im Hinblick auf die Lern- und Leistungsmotivation besonders einflussreiche Erwartung beziehungsweise Überzeugung, die im klassischen Grundmodell unberücksichtigt bleibt, ist die schulische Selbstwirksamkeitserwartung. Diese stellt auch ein Zielkriterium des FOSIS2-Programms dar und wird nachfolgend erläutert.

## **2.2 Schulische Selbstwirksamkeitserwartung – eine Ressource für die Motivation**

*„A strong sense of [self-]efficacy fosters high level of motivation, academic accomplishments, and development of intrinsic interest in academic subject matter“ (Bandura, 1997, S. 174).*

Eine zentrale Ressource für die Motivation und auch für kognitive, emotionale und verhaltensbezogene Prozesse ist die sogenannte *Selbstwirksamkeitserwartung* – ein ursprünglich von Bandura (z. B. 2001) im Rahmen seiner sozial-kognitiven Theorie eingeführtes Konzept, das inzwischen in zahlreichen Kontexten Anwendung gefunden hat und vielfältig beforscht wurde (Gallagher, 2012). Die Selbstwirksamkeitserwartung ist eine persönliche Überzeugung, die im Hinblick auf motivationale und volitionale Prozesse die Intentionsbildung, Zielsetzung, Anstrengung und Ausdauer beim Ausführen einer Handlung beeinflusst (Schwarzer & Jerusalem, 2002). Speziell für die schulische Motivation, für Lernen



und Leistung stellt die Selbstwirksamkeitserwartung somit eine wichtige Ressource dar (Linnenbrink & Pintrich, 2002, 2003; Schunk & Pajares, 2009).

### 2.2.1 Definition und theoretische Konzeption von Selbstwirksamkeitserwartung

Die Selbstwirksamkeitserwartung, synonym auch Kompetenzerwartung genannt (Linnenbrink & Pintrich, 2003), bezeichnet „die subjektive Gewissheit, neue oder schwierige Anforderungssituationen aufgrund eigener Kompetenz bewältigen zu können“ (Schwarzer & Jerusalem, 2002, S. 35). In diesem Sinne beschreibt sie das Vertrauen in die eigenen Bewältigungskompetenzen, schwierige Handlungen nicht nur beginnen, sondern auch erfolgreich beenden zu können (Jerusalem, 2005b; Schwarzer & Jerusalem, 2002). Insbesondere bei Anforderungen, die nicht durch leichte Routinehandlungen bewältigt werden können, sondern die mit einem hohen Grad an Anstrengung und Ausdauer verbunden sind, ist die Selbstwirksamkeitserwartung eine zentrale Ressource für die erfolgreiche Handlungsbe-  
wältigung (Schwarzer & Jerusalem, 2002).

Das Konzept der Selbstwirksamkeitserwartung lässt sich nach den Dimensionen *Spezifität* und *Generalität* sowie *Individualität* und *Kollektivität* differenzieren: Das Vertrauen in die eigenen Bewältigungsfähigkeiten angesichts schwieriger bevorstehender Anforderungen kann sich situationsspezifisch, aber auch situationsübergreifend – im Sinne einer dispositi-  
onalen Überzeugung – äußern: Die sogenannte *situationsspezifische* Selbstwirksamkeit beschreibt die Erwartung, eine konkrete Anforderung erfolgreich bewältigen zu können, auch wenn Barrieren auftreten (z. B. „Ich bin sicher, dass ich den ganzen Abend eisern arbeiten kann, auch wenn andere mich zum Fernsehen einladen“; Schwarzer & Jerusalem, 2002, S. 39). Im Gegensatz dazu umfasst die *allgemeine* Selbstwirksamkeitserwartung sämtliche Lebensbereiche und lässt sich definieren als ein allgemeines, grundsätzliches Vertrauen in die eigenen Lebensbewältigungsfähigkeiten (z. B. „Wenn ein Problem auf-  
taucht, kann ich es aus eigener Kraft meistern“; Schwarzer & Jerusalem, 2002, S. 40). Zwischen situationsspezifischer und allgemeiner Selbstwirksamkeitserwartung sind so-  
genannte *bereichsspezifische* Erwartungen angesiedelt, wie z. B. die *Lehrer-Selbstwirksamkeitserwartung* als eine berufsspezifische personale Ressource oder die *schulische Selbstwirksamkeitserwartung* (Jerusalem, 2005b; Schwarzer & Jerusalem, 2002). Letztere wird auch in der vorliegenden Arbeit untersucht. Sie ist definiert als die subjektive Überzeugung beziehungsweise das Vertrauen des Schülers in die eigenen Fähigkeiten, schwierige schulische Anforderungen bewältigen zu können (z. B. „Ich kann auch die schwierigen Aufgaben im Unterricht lösen, wenn ich mich anstrenge“ Schwarzer & Jerusa-

lem, 2002, S. 40). Das zunächst individuell konzipierte Konstrukt der Selbstwirksamkeitserwartung lässt sich auch kollektiv als eine Art Gruppenüberzeugung fassen. Unter einer kollektiven Selbstwirksamkeitserwartung wird das Vertrauen einer gesamten Gruppe verstanden, schwierige Anforderungen gemeinsam bewältigen zu können (z. B. „Auch mit außergewöhnlichen Vorfällen können wir zurechtkommen, da wir uns im Kollegium Rückhalt bieten“; Jerusalem, 2005b, S. 442).

### **2.2.2 Entstehungsbedingungen von Selbstwirksamkeitserwartung**

Bandura (1997) postuliert vier Quellen, die zur Entstehung von Selbstwirksamkeitserwartungen beitragen: 1. *eigene Erfahrungen*, 2. *stellvertretende Erfahrungen*, 3. *verbale Überzeugungen* und 4. *emotionale und physiologische Erregung*.

Die erste Quelle, die direkte eigene Erfahrung, beeinflusst die Entstehung der Selbstwirksamkeitserwartung am stärksten (Bandura, 1997; Gallagher, 2012). Erlebt eine Person, dass ihre Handlung erfolgreich war und führt diese Erfolgserfahrung auf die eigenen Anstrengungen und Fähigkeiten zurück, so wird ihre Selbstwirksamkeitserwartung in diesem Bereich steigen (Schwarzer & Jerusalem, 2002). Erzielt ein Schüler beispielsweise in naturwissenschaftlichen Fächern gute Noten und interpretiert dies als Folge seines Lernverhaltens sowie seiner naturwissenschaftlichen Fähigkeiten, so werden auch seine naturwissenschaftlichen Kompetenzerwartungen steigen (Usher & Pajares, 2008). Im Gegensatz dazu würde eine kontinuierliche Erfahrung von Misserfolgen, deren Ursache ebenfalls in den eigenen Fähigkeiten verortet wird, die Selbstwirksamkeitserwartung schwächen. Hat sich jedoch erst einmal eine hohe Kompetenzerwartung ausgebildet, so verringert sich auch die negative Wirkung von Misserfolgen. Diese werden dann sogar teilweise als Lernchance interpretiert (Schwarzer & Jerusalem, 2002). Insgesamt hat die Erfahrung eigener Erfolge einen besonders starken Einfluss auf die Ausbildung von Selbstwirksamkeitserwartungen, wenn bei der Bewältigung der Anforderung Widerstände überwunden wurden und die Handlung eigenständig, ohne fremde Hilfe ausgeführt wurde (Usher & Pajares, 2008).

Die zweitstärkste Quelle, die zum Erwerb von Kompetenzerwartungen beiträgt, ist die stellvertretende Erfahrung durch die Beobachtung von erfolgreich handelnden Verhaltensmodellen (Bandura, 1997; Gallagher, 2012; Schwarzer & Jerusalem, 2002). Insbesondere bei noch unbekannten Aufgabengebieten ist das Modelllernen eine wichtige Quelle für die Entwicklung von Selbstwirksamkeitserwartungen (Gallagher, 2012; Usher & Pajares, 2008). Als Verhaltensmodelle eignen sich besonders gut Personen, die dem Lernenden in bedeutsamen Eigenschaften ähneln (z. B. gleiches Alter, Geschlecht, ähnliche Interessen)

(Schwarzer & Jerusalem, 2002; Usher & Pajares, 2008), wobei nach Bandura (1997) eine solche Ähnlichkeit nicht zwangsläufig gegeben sein muss, um Modelllernen zu ermöglichen. Auch „sich selbst enthüllende Bewältigungsmodelle“ (Schwarzer & Jerusalem, 2002, S. 43), sogenannte ‚*Coping Models*‘, die bei der Lösung von Aufgaben Probleme haben und offen mitteilen, wie sie mit diesen Schwierigkeiten umgehen, stellen geeignete Modelle dar (Usher & Pajares, 2008). Zu beachten ist, dass die Beobachtung von Verhaltensmodellen natürlich auch zu einer Verringerung der Selbstwirksamkeitserwartung beitragen kann, wenn z. B. die Modelle bei der Lösung eines Problems, insbesondere bei der Bewältigung einfacher Aufgaben, scheitern (Schunk & Pajares, 2009; Usher & Pajares, 2008).

Etwas schwächer als die ersten beiden Quellen trägt auch die sogenannte verbale Überzeugung zum Aufbau von Kompetenzerwartungen bei (Bandura, 1997; Gallagher, 2012). Hierzu zählen verbale Ermutigungen, Zuspruch und Leistungsrückmeldungen, Lob sowie allgemein ein geäußertes Vertrauen in die Kompetenzen einer anderen Person (Schwarzer & Jerusalem, 2002). Im Kontext Schule sind es speziell die Rückmeldungen und Ermutigungen durch Lehrer, Eltern und Peers, die die schulische Selbstwirksamkeitserwartung verbessern. Insbesondere dann, wenn die Schüler noch nicht über ausreichende Fähigkeiten verfügen, sich selbst einzuschätzen und sie noch stark von der Bewertung anderer abhängen, sind solche Ermutigungen wichtig (Usher & Pajares, 2008). Die Selbstwirksamkeitserwartung kann jedoch nicht durch jede Form von Ermutigung oder Art von Feedback verbessert werden (Schunk & Pajares, 2009) (förderliche Bedingungen und Merkmale von günstigen Lern- und Leistungsrückmeldungen werden im Abschnitt 2.2.6 beschrieben).

Die emotionale und physiologische Erregung ist die vierte und schwächste Quelle der Selbstwirksamkeitserwartung (Bandura, 1997; Gallagher, 2012). Gemeint sind körperliche Reaktionen und Emotionen, die vor, während und nach der Bewältigung einer Anforderung auftreten können (Margolis & McCabe, 2006). So haben in einer bestimmten Situation z. B. auch Ängstlichkeit, Stresserleben, Müdigkeit sowie die generelle Befindlichkeit einen Einfluss darauf, wie das eigene Verhalten bewertet wird (Gallagher, 2012; Schunk & Pajares, 2009; Usher & Pajares, 2008). Das Erleben von Ängstlichkeit und Aufgeregtheit in der Schule kann von dem Betroffenen z. B. als Beleg für einen Mangel an Fähigkeiten interpretiert werden und somit zu einer Reduktion der Kompetenzerwartungen führen (Schwarzer & Jerusalem, 2002; Usher & Pajares, 2008).

### **2.2.3 Auswirkungen und Korrelate von Selbstwirksamkeitserwartung**

Da die Selbstwirksamkeitserwartung viele kognitive, motivationale, emotionale und verhaltensbezogene Prozesse beeinflusst, stellt sie für zahlreiche Bereiche und psychologische Subdisziplinen eine relevante personale Ressource dar. So zeigen sich beispielsweise bedeutsame Zusammenhänge zwischen der Selbstwirksamkeitserwartung und Gesundheitsbeziehungsweise Risikoverhaltensweisen (z. B. Rauchen), zwischen der Selbstwirksamkeitserwartung und klinisch-psychologischen Prozessen (z. B. depressive Symptome) und auch für die Bereiche Sport, Beruf und Karriere hat die Selbstwirksamkeitserwartung eine große Bedeutung. Insgesamt verfügen hoch selbstwirksame Personen über bessere Stress- und Lebensbewältigungsfähigkeiten als niedrig selbstwirksame (Jerusalem, 2005b; Schwarzer & Jerusalem, 2002) und über alle Anwendungsfelder hinweg hat sich das Konstrukt als ein guter Prädiktor für Wohlbefinden und Erfolg erwiesen (Bandura, 1997; Gallagher, 2012; Linnenbrink & Pintrich, 2002).

#### ***Auswirkungen auf Motivation, Lernen und Leistung***

Es wurde bereits deutlich, dass der Selbstwirksamkeitserwartung in der Motivationspsychologie eine bedeutende Rolle zukommt. Bei gleicher Fähigkeit begegnen Menschen mit hoher Selbstwirksamkeitserwartung Anforderungen zuversichtlicher, nutzen bessere Bewältigungsstrategien, lassen sich durch Rückschläge weniger entmutigen, werten Fortschritte eher als Belege für die eigene Kompetenz, sind motivierter sich anzustrengen und haben ein höheres Wohlbefinden als Menschen mit niedrigerer Selbstwirksamkeitserwartung. Die Selbstwirksamkeitserwartung trägt somit insgesamt zu einer besseren Selbstregulation bei (Bandura, 1997; Jerusalem, 2005b; Lee, Lee & Bong, 2014; Linnenbrink & Pintrich, 2002; Schunk & Pajares, 2009; Schwarzer & Jerusalem, 2002; Wang, Kim, Bong & Ahn, 2013). Usher und Pajares (2008) weisen zudem darauf hin, dass die Selbstwirksamkeitserwartung auch mit anderen motivational relevanten Einflussgrößen zusammenhängt, z. B. mit der Kausalattribution, dem Selbstkonzept der Begabung, dem Optimismus, der motivationalen Orientierungen, der Ängstlichkeit und Wertzuschreibungen.

Die motivationalen Auswirkungen der Selbstwirksamkeitserwartung lassen sich auch auf den Kontext Schule übertragen: Bei gleicher Fähigkeit weisen Schüler mit höherer schulischer Selbstwirksamkeitserwartung günstigere Voraussetzungen zur Leistungserbringung auf als Schüler mit niedriger Selbstwirksamkeitserwartung: Hoch selbstwirksame Schüler strengen sich mehr an, verfolgen Aufgaben mit größerer Ausdauer, setzen sich höhere Ansprüche, zeigen eine größere Flexibilität und Kreativität bei der Problemlösung, sind effek-

tiver im Zeitmanagement und nehmen realistischere Bewertungen ihrer Leistungen sowie vermehrt selbstwertschützende Kausalattributionen vor (Bandura, 1997; Jerusalem, 2005b; Schunk & Pajares, 2009; Schwarzer & Jerusalem, 2002; Usher & Pajares, 2008). In welchem Maße Schüler in schwierigen Situationen beziehungsweise angesichts von hohen Anforderungen mit Ängstlichkeit oder Gelassenheit reagieren, hängt ebenfalls maßgeblich von ihren Selbstwirksamkeitserwartungen ab (Usher & Pajares, 2008). Schüler mit hohen Kompetenzerwartungen zeigen generell mehr positive Emotionen, wie z. B. Zufriedenheit und Stolz (Linnenbrink & Pintrich, 2003). Zudem weisen Schüler mit hohen schulischen Selbstwirksamkeitserwartungen auch höhere soziale Kompetenzen auf und scheinen stärker bestrebt, soziale Verantwortung zu übernehmen (Linnenbrink & Pintrich, 2002). Darüber hinaus zeigen Studien die Selbstwirksamkeitserwartung als einen sehr guten Prädiktor für akademische Leistungen auf (Lee et al., 2014; Linnenbrink & Pintrich, 2002; Schunk & Pajares, 2009; Usher & Pajares, 2008; Wang et al., 2013; Wright, Jenkins-Guarnieri & Murdock, 2013). Anders als hochwirksame Schüler fallen niedrig selbstwirksame nicht selten durch Motivationsprobleme auf: Da sie überzeugt sind, ihre eigenen Kompetenzen würden nicht genügen, um Erfolge zu erzielen, wenden sie vermehrt Vermeidungs- und Fluchtverhaltensweisen an, geben schneller auf und unternehmen eher oberflächliche Bewältigungsversuche. Dadurch erleben niedrig selbstwirksame Schüler häufiger Misserfolge, die sie im Sinne einer sich selbsterfüllenden Prophezeiung wiederum als Belege für ihre mangelnden Fähigkeiten verstehen, wodurch ihre Selbstwirksamkeitserwartung weiterhin niedrig bleibt (Margolis & McCabe, 2006) und sich das Vermeidungsverhalten im Sinne der erlernten Hilflosigkeit von Seligman (1975) manifestieren kann. Schüler mit einer eher gering ausgeprägten Kompetenzüberzeugung erleben auch vermehrt negative Emotionen in der Schule, wie z. B. Angst, Frustration und Ärger (Linnenbrink & Pintrich, 2003). Lernschwierigkeiten, schlechtere Leistungen, Schulprobleme und Konflikte mit den Lehrkräften treten ebenfalls häufiger bei Schülern auf, die in mehreren schulischen Bereichen eine niedrige Selbstwirksamkeitserwartung aufweisen als bei hoch wirksamen Schülern (Margolis & McCabe, 2006).

Linnenbrink & Pintrich (2003) fassen die motivationalen Auswirkungen einer hohen Selbstwirksamkeitserwartung auf Lernen und Leistung im Kontext Schule folgendermaßen zusammen: Eine hohe schulische Selbstwirksamkeitserwartung steigert das *verhaltensbezogene*, das *kognitive* und das *motivationale Engagement* im Klassenraum. Unter *verhaltensbezogenem Engagement* verstehen die Autoren beobachtbare Verhaltensweisen wie Anstrengungs- und Ausdauerbemühungen oder auch die Suche nach Unterstützung. Nicht

jedes Aufsuchen von Hilfe ist jedoch funktional. Manchmal kann hilfesuchendes Verhalten auch Ausdruck von Lernvermeidung sein. Dient die Suche nach Hilfe bei Lehrern und Peers aber wirklich dem Lernverständnis, wird diese als instrumentelle, funktionale oder adaptive Hilfesuche bezeichnet. Studien belegen den Zusammenhang von hoher Selbstwirksamkeitserwartung und einer solchen funktionalen Suche nach Hilfe. Niedrig selbstwirksame Personen hingegen interpretieren die eigene Suche nach Unterstützung oft als Beleg für ihre mangelnden Fähigkeiten, verringern somit ihr hilfesuchendes Verhalten und reagieren mit Hilflosigkeit. Eine hohe Selbstwirksamkeitserwartung fördert auch das *kognitive Engagement* der Schüler. Hierzu zählen der Einsatz von Lernstrategien und Metakognitionen. Studien belegen einen positiven Zusammenhang von Selbstwirksamkeitserwartung und dem Einsatz von Tiefen-Lernstrategien wie Elaboration. Auch metakognitive Strategien, wie z. B. die Reflexion und die Regulation sowie die Evaluation des eigenen Lernens, gehen mit höherer Selbstwirksamkeitserwartung und Performanz einher. Die Selbstwirksamkeitserwartung wirkt sich zudem auf das sogenannte *motivationale Engagement* im Klassenraum aus und führt zu einem Zugewinn an Interesse, positiven Lernemotionen sowie zu einer für den Schüler wachsenden Bedeutung des Lerngegenstands. Die Beziehungen zwischen den einzelnen Variablen sind reziprok. Das heißt, eine hohe Selbstwirksamkeitserwartung führt zu einem höheren Engagement, so wie umgekehrt auch ein verstärktes Engagement die eigenen Selbstwirksamkeitserwartungen erhöhen kann (Linnenbrink & Pintrich, 2003).

All die geschilderten Zusammenhänge könnten die Vermutung nahelegen, die Selbstwirksamkeitserwartung sollte so hoch wie möglich ausgeprägt sein. Es zeigt sich jedoch, dass nur realistische, im Hinblick auf die tatsächlichen eigenen Fähigkeiten angemessene beziehungsweise minimal höhere Selbstwirksamkeitserwartungen die oben geschilderten positiven Auswirkungen haben (Bandura, 1997; Pintrich, 2003). Zu hohe Selbstwirksamkeitserwartungen, die auf einer deutlichen Überschätzung der tatsächlichen Kompetenzen basieren, können hingegen zu einer Verringerung des Lernverhaltens beitragen (Linnenbrink & Pintrich, 2003; Schunk & Pajares, 2009): Wenn Schüler der Überzeugung sind, sie könnten bereits alles, zeigen sie möglicherweise nur noch geringe oder gar keine Anstrengungsleistungen.

#### 2.2.4 Mit Selbstwirksamkeitserwartung verwandte Konstrukte

In der Psychologie werden auch andere Überzeugungen und Erwartungen beschrieben, die dem Konzept der Selbstwirksamkeitserwartung ähneln und doch davon unterschieden werden können. Bandura (2001) selbst weist in seiner sozial-kognitiven Theorie auf einen zweiten Typ von Erwartungen hin, die *Ergebnis-* oder *Konsequenzerwartungen*. Diese beziehen sich auf Handlungsergebnisse und setzen bestimmte Handlungen mit spezifischen Resultaten in Beziehung (Jerusalem, 2005b). Ein Schüler könnte beispielsweise erwarten, dass die intensiven Lernvorbereitungen für eine Klausur ein bestimmtes Notenergebnis nach sich ziehen (Schwarzer & Jerusalem, 2002). Anders als die Selbstwirksamkeitserwartungen weisen Ergebniserwartungen nicht notwendigerweise einen Selbstbezug auf, sondern können zunächst auch ganz allgemeine Erwartungen beschreiben („Wenn man das geforderte Fächerwissen beherrschen lernt, besteht man die Prüfung“, Schwarzer & Jerusalem, 2002, S. 36). Ein übergeordnetes Modell, das verschiedene Erwartungen und auch Wertvorstellungen integriert und vielfach im Bereich der Lern- und Leistungsmotivation Anwendung gefunden hat, ist das sogenannte *erweiterte kognitive Motivationsmodell*. Das Modell wurde von Heckhausen (1977) in Anlehnung an Forschungsarbeiten von Vroom (1964) vorgeschlagen. Hiernach sind für die Ausprägung der Motivation im Handlungsverlauf *Situations-Ergebnis-*, *Handlungs-Ergebnis-* und *Ergebnis-Folge-Erwartungen* sowie außerdem *Folgen-* und *Tätigkeitsanreize* entscheidend. Die Situations-Ergebnis-Erwartung beinhaltet Annahmen darüber, inwieweit ein Ergebnis bereits durch eine Situation festgelegt erscheint. Ein Schüler, der sich seines Könnens sicher ist, wird davon ausgehen, dass er eine Prüfung auch ohne weitere zusätzliche Lernbemühungen erfolgreich meistern kann. Eine hohe Situations-Ergebnis-Erwartung verringert folglich die Lernanstrengungen. Handlungs-Ergebnis-Erwartungen umfassen Überzeugungen einer Person darüber, wie wahrscheinlich bestimmte selbst ausgeübte Handlungen ein spezifisches Ergebnis nach sich ziehen. Ist ein Schüler beispielsweise davon überzeugt, dass der Ausgang einer Klausur von seinen eigenen Lernbemühungen abhängt, so hat er eine hohe Handlungs-Ergebnis-Erwartung, die dann auch die Lernhandlungen verstärken wird. Ergebnis-Folge-Erwartungen sind Vermutungen darüber, inwieweit ein Ergebnis mit spezifischen Folgen verbunden ist (Rheinberg, 2010). Die Annahme aufgrund einer sehr guten Abschlussnote einen gewünschten Beruf erlernen zu können, wäre eine solche Ergebnis-Folge-Erwartung (Rheinberg & Vollmeyer, 2012). Fallen diese Erwartungen und der Folgenanreiz hoch aus – Letzteres ist dann der Fall, wenn die Folgen besonders attraktiv erscheinen – werden die Handlungstendenzen erhöht (Rheinberg, 2010). Etwas später wurde das erweiterte kogni-

ve Modell noch um die sogenannten tätigkeitsspezifischen Vollzugsanreize ergänzt (Rheinberg, 1989). Diese beinhalten die Anreize, die in der Ausübung einer Tätigkeit selbst verankert sind, sodass diese „um ihrer selbst willen“ (Rheinberg, 2010, S. 375) betrieben wird. Nach Rheinberg und Vollmeyer (2012) ist die Selbstwirksamkeitserwartung in dem erweiterten kognitiven Motivationsmodell in der Handlungs-Ergebnis-Erwartung verborgen. Diese fasst nämlich nicht nur die bereits von Bandura (1977; 2001) beschriebene Ergebniserwartung, nach der eine Handlung ein bestimmtes Ergebnis zur Folge hat, sondern sie beinhaltet zugleich das Zutrauen in die eigenen Fähigkeiten, diese Handlung auch ausführen zu können.

Auf die Bedeutsamkeit des eigenen Kompetenzerlebens weisen auch andere psychologische Theorien hin. So postulieren beispielsweise Deci und Ryan (1993) im Rahmen ihrer *Theorie der Selbstbestimmung* drei dem Menschen innewohnende Grundbedürfnisse, deren Erfüllung Antriebsquelle für motiviertes Handeln darstellt: das Bedürfnis nach *Selbstbestimmung*, das Bedürfnis nach *sozialer Eingebundenheit* und schließlich das Bedürfnis nach *Kompetenzerfahrung*. Letzteres beschreibt das Bedürfnis (nicht die Erwartung), selbst etwas bewirken und Anforderungen bewältigen zu können (Krapp & Ryan, 2002). Heider hat bereits im Jahr 1958 darauf hingewiesen, dass der Mensch dazu neigt, Ursachenzuschreibungen für Handlungsausgänge vorzunehmen. Diese sogenannten naiven *Kausalattributionen* einer Person wurden von Weiner und Kollegen (1971; 1986) auf den Leistungskontext übertragen. Nach den Autoren hängt die emotionale Bewertung von Erfolgen oder Misserfolgen wesentlich davon ab, inwiefern eine Person die Ursachen des Handlungsausgangs bei sich selbst sieht und davon überzeugt ist, dass die eigenen Fähigkeiten oder Anstrengungsbemühungen zu dem Ergebnis beigetragen haben. Die Frage nach den Faktoren, durch die Handlungsergebnisse beeinflusst werden, wird auch mit dem Konzept der *Kontrollüberzeugungen* aufgegriffen (Rotter, 1966). Sieht eine Person sich selbst als Verursacher ihrer Handlungen, so verfügt sie über *internale Kontrollüberzeugungen*. Glaubt sie hingegen, dass andere äußere Faktoren, wie z. B. der Zufall oder das Glück, für die Ergebnisse einer Handlung verantwortlich sind, so hat sie eine *externale Kontrollüberzeugung*, da der Ort der Handlungskontrolle extern und nicht intern verortet wird (Reimann & Hammelstein, 2006). Auch wenn die Konzepte der Selbstwirksamkeitserwartung und der internalen Kontrollüberzeugungen ähnlich klingen, handelt es sich empirisch nachweislich um unterschiedliche Konzepte. Die Kontrollüberzeugungen beziehen sich auf das Handlungsergebnis und sind, ähnlich wie auch die Kausalattributionen, als Ursachenerklärungen für Handlungsausgänge zu fassen.



Hingegen beschreibt die Selbstwirksamkeitserwartung das Vertrauen in die eigenen Fähigkeiten, Handlungen auch ausführen zu können (Gallagher, 2012; Maddux, 1995).

Als mit der Selbstwirksamkeitserwartung ebenfalls verwandte Konstrukte sind auch der *dispositionale Optimismus* (Scheier & Carver, 1987) und der *Kohärenzsinn* (Antonovsky, 1993) zu nennen. Unter dem dispositionalen Optimismus wird die generalisierte Überzeugung beziehungsweise allgemeine Erwartung verstanden, dass die Dinge grundsätzlich eher gut verlaufen werden (Carver, Scheier & Segerstrom, 2010). Wie auch schon die Ergebniserwartung enthält auch der Optimismus nicht zwangsläufig einen Selbstbezug und lässt im Gegensatz zur Selbstwirksamkeitserwartung offen, ob einem gute Dinge aufgrund eigenen Zutuns oder anderer äußerer Faktoren widerfahren (Gallagher, 2012; Reimann & Hammelstein, 2006). Der Kohärenzsinn bezeichnet ein globales Lebensgefühl, das die Umwelt und das Leben im Allgemeinen als verstehbar, bewältigbar und bedeutsam erscheinen lässt (Reimann & Hammelstein, 2006). Kohärenzsinn und Selbstwirksamkeitserwartung sind miteinander korreliert, wobei der Kohärenzsinn als ein übergeordnetes, allgemeineres Konstrukt verstanden werden kann (Kröninger-Jungaberle & Grevenstein, 2013).

Weitere Konzepte, die enge Zusammenhänge mit der Selbstwirksamkeitserwartung aufweisen, aber theoretisch und empirisch von dieser getrennt werden können, sind das Selbstkonzept und der Selbstwert. Das Selbstkonzept umfasst generellere Überzeugungen bezüglich der eigenen Fähigkeiten (z. B. „I’m good at math“; Linnenbrink & Pintrich, 2003, S. 121), die auf der Basis retrospektiver Erfahrungen entwickelt wurden. Im Gegensatz dazu beinhalten Selbstwirksamkeitserwartungen prospektive, auf ein Ziel in der Zukunft gerichtete Überzeugungen über die eigenen Bewältigungsfähigkeiten angesichts von Schwierigkeiten („I’m still confident I can solve these quadratic equations today, even though the last time I had algebra was over 30 years ago“; Linnenbrink & Pintrich, 2003, S. 121). Studien belegen, dass sich die Selbstwirksamkeitserwartung und das Selbstkonzept auch empirisch trennen lassen (Ferla, Valcke & Cai, 2009) und die Selbstwirksamkeitserwartungen eine größere Vorhersagekraft für Motivation und Lernverhalten haben als das allgemeinere Selbstkonzept (Linnenbrink & Pintrich, 2003). Der Selbstwert kann als die emotionale Tönung des Selbstkonzepts verstanden werden, wie z. B. Stolz oder Scham bezüglich der eigenen (mangelnden) Fähigkeiten (Linnenbrink & Pintrich, 2003; Schunk & Pajares, 2009).

### **2.2.5 Erfassung von Selbstwirksamkeitserwartung**

Das Konzept der Selbstwirksamkeitserwartung wurde bisher in zahlreichen Studien auf sehr unterschiedliche Weise erfasst (z. B. Bong, 2004; Ferla et al., 2009; Klassen, 2002; Kröninger-Jungaberle & Grevenstein, 2013; Lee et al., 2014; Pajares, 1996; Wang et al., 2013). Vielen Instrumenten wird eine unzureichende und unangemessene Operationalisierung der Selbstwirksamkeitserwartung vorgeworfen (Pajares, 1996; Schunk & Pajares, 2009). Bereits Bandura (1986; 1997) hat darauf hingewiesen, dass im Sinne der theoretischen Konzeption der Selbstwirksamkeitserwartung bei der Erfassung des Konstrukts unbedingt der Anwendungsbereich berücksichtigt und die Selbstwirksamkeitserwartung bezogen auf schwierige oder herausfordernde Aufgaben gemessen werden sollte. Nicht selten wird diese Forderung jedoch missachtet und die Selbstwirksamkeitserwartung ohne Bezug auf eine Schwierigkeit oder ein Hindernis erfasst (Pajares, 1996; Schunk & Pajares, 2009), sodass eine Abgrenzung zu verwandten Konzepten wie dem Selbstkonzept, dem Selbstwert (Schunk & Pajares, 2009) oder auch der Ergebniserwartung (Schunk, 1995b) nicht mehr eindeutig möglich ist. In der vorliegenden Untersuchung werden diese Warnungen berücksichtigt und die von Jerusalem und Satow (1999) entwickelte bereichsspezifische Skala „Schulische Selbstwirksamkeitserwartung“ verwendet. Eine genaue Beschreibung der Skala findet sich in Unterkapitel 6.4.1.

### **2.2.6 Implikationen für die Praxis: Fördermöglichkeiten von Selbstwirksamkeitserwartung**

Immer wieder stellen sich Lehrkräfte die Frage, wie sie ihre Schüler, insbesondere leistungsschwache Schüler, die wenig Zutrauen in ihre eigenen Fähigkeiten besitzen (Margolis & McCabe, 2006), motivieren und fördern können (Fries & Souvignier, 2009; Rheinberg & Krug, 2005b). Das Konzept der Selbstwirksamkeitserwartung als ein wesentlicher Stabilisator für Lern- und Leistungsmotivation (Linnenbrink & Pintrich, 2003) bietet durch die Quellen zur Entstehung der Selbstwirksamkeitserwartung eine hilfreiche theoretische Grundlage, aus der praktische Hilfen zur Beantwortung dieser Frage abgeleitet werden können (Margolis & McCabe, 2006; Schunk & Pajares, 2009). Eine Schule, die ihren Schülern viele Gelegenheiten bietet, eigene und stellvertretende Erfolge sowie verbale Ermutigung und positive Gefühle zu erfahren, müsste nach der Theorie der Selbstwirksamkeitserwartung zu einer Stärkung der eigenen Kompetenzüberzeugungen beitragen (Bandura, 1997).

### ***Eigene Erfahrungen***

In Bezug auf die erste Quelle sollte der Unterricht den Schülern viele Angebote für die Erfahrung eigener Erfolge ermöglichen. Dafür sind vor allem herausfordernde Aufgaben mittleren Schwierigkeitsgrads geeignet. Denn die erfolgreiche Bewältigung minimal herausfordernden Anforderungen wird in der Regel positiver bewertet als die Erledigung sehr leichter Aufgaben (Jerusalem, 2005b; Linnenbrink & Pintrich, 2002, 2003; Margolis & McCabe, 2006). Schulische Erfolgserfahrungen sind leichter möglich, wenn eine Fehlerkultur im Unterricht vorherrscht (Wigfield et al., 2012) und das Setzen von Nahzielen beziehungsweise die Unterteilung von großen, langfristigen Lernzielen in kleinere, kurzfristig zu erreichende Teilziele unterstützt wird (Jerusalem, 2005b; Schiefele, 2009; Schunk, 1995a). Insbesondere selbst gesetzte Ziele, die sich auf spezifische, herausfordernde und naheliegende Aufgaben beziehen, führen zu verstärkten Lernhandlungen und sind besonders geeignet, die Selbstwirksamkeitserwartung zu erhöhen (Schunk, 1995a). Schunk (1995a) nimmt außerdem an, dass Ziele, die den Handlungsverlauf beziehungsweise den Bewältigungsprozess (Prozessziele) und nicht das Ergebnis einer Handlung (Produktziel) betreffen, ebenfalls die Ausbildung der Selbstwirksamkeitserwartung fördern. Auch ein binnendifferenzierter (Wigfield et al., 2011) und autonomie-förderlicher Unterricht, der Schülern Wahlfreiheiten und Mitbestimmungsmöglichkeiten anbietet, kann Selbstwirksamkeitserwartung und Lernmotivation steigern (Schiefele, 2009; Schunk & Pajares, 2009; Wigfield et al., 2011). Denn die Schüler werden dadurch in ihrer Eigeninitiative und ihrem schulischen Kontrollerleben gestärkt, und so entstehen wiederum mehr Gelegenheiten für eigene Erfolge (Schiefele, 2009; Wigfield et al., 2011). Neben Maßnahmen zur Autonomieförderung kann auch der Einbezug des Interesses der Schüler die Aufmerksamkeit, die Anstrengungsbereitschaft und damit ebenfalls die Unterrichtsgelegenheiten für eigene Erfolge erhöhen (Margolis & McCabe, 2006).<sup>5</sup> Auch durch das explizite Unterrichten und Anleiten in der angemessenen und richtigen Anwendung von Lernstrategien (z. B. Wiederholen und Strukturieren des Lernstoffs) (Margolis & McCabe, 2006) und Selbstregulationsfähigkeiten (Lavasani, Mirhosseini, Hejazi & Davoodi, 2011) werden vermehrt Lernerfolge erzielt. Der Lehrer sollte sich dabei auf einzelne Strategien beschränken und die Ein-

---

<sup>5</sup> Zur Förderung des Interesses ist es nach Margolis und McCabe (2006) besonders hilfreich, das Neue und Unbekannte sowie die Bedeutsamkeit des Lerngegenstands bezogen auf die Lebensrealität der Schüler hervorzuheben. Im Sinne des ‚Catch- and Hold‘-Prinzips (Interesse wecken und aufrechterhalten) schafft das Neue und Unbekannte zunächst Neugier und ermöglicht somit Aufmerksamkeit. Das Hervorheben der Relevanz eines Themas hilft, das Interesse längerfristig aufrechtzuerhalten (Margolis & McCabe, 2006). Die Bedeutsamkeit des Lerngegenstands kann z. B. durch das Ausformulieren persönlich bedeutender Lernziele oder das Hervorheben praktischer Anwendungsmöglichkeiten des Themas verdeutlicht werden (Schiefele, 2009).

übung an verschiedenen Aufgabentypen erproben und optimieren sowie die Internalisierung der Lernstrategien beim Schüler sicherstellen (Margolis & McCabe, 2006). Eine solche Internalisierung der Lernstrategien kann durch die Anleitung zur Selbstverbalisation beziehungsweise Selbstinstruktion sowie zur Selbstbeobachtung und Selbstevaluation unterstützt werden (Margolis & McCabe, 2006; Schunk & Pajares, 2009).

### ***Stellvertretende Erfahrungen***

Für eine Stärkung der Selbstwirksamkeitserwartung durch die Nutzung der zweiten Quelle, der stellvertretenden Erfahrungen, sind Lehrer und Eltern, insbesondere aber auch Peers als Modelle geeignet (Schunk & Pajares, 2009). Das Modelllernen gelingt besonders gut, wenn die Modelle Ähnlichkeiten zum Lernenden aufweisen z. B. in Bezug auf das Alter, die ethnische Herkunft, das Geschlecht, die Fähigkeiten, Interessen, Kleidung, das soziale Umfeld und das Leistungsniveau (Margolis & McCabe, 2006). Grundsätzlich wird unterschieden zwischen ‚Mastery-‘ und ‚Coping Models‘. Erstere demonstrieren bereits vorhandene Leistungen und Fähigkeiten, wohingegen Letztere den Weg der Bewältigung und Überwindung von Schwierigkeiten und Hindernissen aufzeigen. Niedrig selbstwirksame Schüler mit Lernschwierigkeiten profitieren besonders von den ‚Coping Models‘. Im Unterricht bietet das gemeinsame Arbeiten in kooperativen Lerngruppen einen Rahmen, in dem die Beobachtung und Nachahmung erfolgreicher Verhaltensmodelle stellvertretende Erfahrungen begünstigen (Jerusalem, 2005a). Kooperative Lernformen, die eine leistungsheterogene Zusammensetzung der Gruppen, klar strukturierte Aufgabeninstruktionen und eine positive Abhängigkeit der Gruppenmitglieder untereinander vorsehen, haben sich auch als hilfreich für die Ausbildung sozialer Kompetenzen erwiesen (Wigfield et al., 2011; Wigfield et al., 2012). Deren Erwerb sowie das Erleben von gelungenen sozialen Interaktionen können ebenfalls als Erfolgserfahrung gewertet werden und in diesem Sinne zu einer Stärkung der Selbstwirksamkeitserwartung beitragen (Felner, Seitinger, Brand, Burns & Bolton, 2007; Hudley, Graham & Taylor, 2007).

### ***Verbale Überzeugung***

Von der dritten Quelle der Selbstwirksamkeitserwartung lassen sich Empfehlungen für die Schüler-Lehrer-Kommunikation ableiten. So sollten die Lehrkräfte ihre Schüler in dem Glauben bestärken, dass sie mit einer gewissen angemessenen Anstrengung und Ausdauer in der Lage sind, die anstehenden Lernaufgaben zu bewältigen und Erfolge zu erzielen (Linnenbrink & Pintrich, 2003; Margolis & McCabe, 2006). Zu beachten ist dabei, dass

unglaubliche Rückmeldungen die Kompetenzüberzeugungen verringern können (Usher & Pajares, 2008). Förderliche Rückmeldungen und Ermutigungen sollten deshalb realistisch und glaubwürdig sein (Margolis & McCabe, 2006; Usher & Pajares, 2008) und vor allem bereichsspezifische und nicht globale, allgemeine Selbstwirksamkeitserwartungen stärken. Um Lösungswege aufzuzeigen und eine Grundlage für zukünftige Erfolgserfahrungen zu schaffen, sollte vor allem korrekatives Feedback eingesetzt werden, das Hinweise auf Korrektur- und Verbesserungsmöglichkeiten enthält (Margolis & McCabe, 2006). Darüber hinaus sollte den Rückmeldungen vor allem die individuelle Bezugsnorm zugrunde liegen, sodass die personale Entwicklung und nicht der soziale Vergleich hervorgehoben wird (Usher & Pajares, 2008; Wigfield et al., 2011; Wigfield et al., 2012). Durch solche Rückmeldungen, die den individuellen Lernfortschritt aufzeigen, nehmen die Schüler ihre eigenen (Teil-)Erfolge eher wahr (Wigfield et al., 2012). Besonders hilfreich für den Aufbau von Kompetenzerwartungen haben sich darüber hinaus Rückmeldungen erwiesen, die nicht nur Bewertungen, sondern auch Ursachen von Handlungsergebnissen enthalten (*attributionales Feedback*) (Jerusalem, 2005b; Siegle & McCoach, 2007). Ein förderliches attributionales Feedback führt Erfolge auf die Fähigkeiten und Anstrengungsbemühungen der Schüler zurück (Schiefele, 2009; Wigfield & Wentzel, 2007) und begründet Misserfolge mit variablen, veränderbaren Ursachen wie z. B. mit mangelnden Lernhandlungen und einer zu gering ausgeprägten Anstrengungsbereitschaft (Linnenbrink & Pintrich, 2003; Margolis & McCabe, 2006). Betrachtet ein Lehrer die sehr gute Leistung eines Schülers als Folge seiner Fähigkeiten und seiner persönlichen Anstrengung, stärkt dies beim Schüler die Überzeugung, dass er in ähnlichen Bereichen auch künftig erfolgreich sein kann. Wird ein Misserfolg seitens der Lehrkraft auf die mangelnden Lernbemühungen des Schülers zurückgeführt, so wird die Selbstwirksamkeitserwartung des Schülers ebenfalls aufrechterhalten und er ist möglicherweise motiviert, sein Lernen in Zukunft zu intensivieren und sich verstärkt anzustrengen (Jerusalem, 2005b). Zu beachten sind hierbei allerdings sogenannte paradoxe Effekte von Lob und Tadel: Attribuiert der Lehrer beispielsweise eine gute Note bei einem Schüler auf die Fähigkeit und die gleiche Note bei einem anderen auf dessen Anstrengung, so wird Letzterer die Rückmeldung nicht unbedingt als Lob werten, sondern diese möglicherweise als Beleg für seine Unfähigkeit in Erinnerung behalten (Schwarzer & Jerusalem, 2002). Der Einsatz von Lob kann bei all diesen Rückmeldungen überhaupt nur angemessen erfolgen, wenn der Lernende dieses auch wirklich verdient hat. Unbegründetes und ungerechtfertigtes Lob wird vom Schüler schnell als solches entlarvt,

wodurch die Lehrkraft an Glaubwürdigkeit verlieren kann und zukünftiges Lob von vornherein nicht mehr ernst genommen wird (Margolis & McCabe, 2006).

### ***Emotionale und physiologische Erregung***

Auch aus der vierten Quelle der Selbstwirksamkeitserwartung lassen sich bestimmte Förderstrategien ableiten. Das Erleben von Angst und Aufgeregtheit in der Schule kann beispielsweise als Beleg für einen Mangel an Fähigkeiten interpretiert werden und zu einer Reduktion der Kompetenzerwartungen führen (Schwarzer & Jerusalem, 2002; Usher & Pajares, 2008). Eine Reduktion von Stresserleben und auch von Prüfungsangst z. B. durch mehr Transparenz kann somit auch eine günstige Wirkung auf die schulische Selbstwirksamkeitserwartung nach sich ziehen (Röder, 2008). Im Unterkapitel 2.3.6 wird eine Maßnahme zur Förderung von Transparenz, der Einsatz von Transparenzpapieren, ausführlich dargestellt. Auch ein positives Klassenklima sowie eine hohe Zufriedenheit mit dem Unterricht tragen zur Stärkung der schulischen Selbstwirksamkeitserwartung bei (Kolbe, Jerusalem & Mittag, 1998; Satow, 1999). Die Kompetenzerwartung wird ebenfalls durch emotionale Merkmale der Lehrkraft beeinflusst. So fördert eine Lehrperson, die Freude am Unterrichten hat und Begeisterungsfähigkeit zeigt, die Zufriedenheit, das schulische Wohlbefinden und somit auch die Kompetenzüberzeugung ihrer Schüler (Zhang, 2014).

### **2.2.7 Zusammenfassung: Selbstwirksamkeitserwartung**

Ein Zielkriterium, das zur Evaluation des FOSIS2-Programms herangezogen wird, ist die Selbstwirksamkeitserwartung, wobei hier die bereichsspezifische, schulische Selbstwirksamkeitserwartung untersucht wird. Diese beschreibt die subjektive Erwartung, schwierige schulische Anforderungen bewältigen zu können (Schwarzer & Jerusalem, 2002) und ist eine bedeutsame Ressource für die schulische Motivation. So geht sie z. B. mit einer höheren Anstrengung und Ausdauer, höheren Ansprüchen und einer flexibleren Problemlösung einher (z. B. Bandura, 1997; Jerusalem, 2005b). Wichtige Quellen zum Aufbau von Selbstwirksamkeitserwartungen sind eigene und stellvertretende (Erfolgs-)Erfahrungen sowie verbale Überzeugungen. Auch die Interpretation eigener emotionaler und physiologischer Erregung kann zum Aufbau von Selbstwirksamkeitserwartungen beitragen (Bandura, 1997). In diesem Sinne wurden im Rahmen des FOSIS2-Projekts Unterrichtsstrategien entwickelt und zusammengestellt, die darauf abzielen, den Schülern im Unterricht vermehrt Gelegenheiten für die Erfahrung eigener und stellvertretender Erfolge sowie verbaler

Ermutigung und positiver Emotionen anzubieten und somit ihre schulische Selbstwirksamkeitserwartung zu stärken.

In der Motivationsforschung wird gefordert, nicht nur positive Einflussfaktoren und Ressourcen der Motivation zu untersuchen, sondern auch Risikofaktoren zu berücksichtigen, die eine bereits vorhandene Motivation verringern oder den Aufbau einer Motivation verhindern können (Martin, 2008; Pintrich, 2003). Aus diesem Grund werden in der vorliegenden Arbeit auch die Auswirkungen der FOSIS2-Maßnahmen auf Prüfungsängstlichkeit und Stresserleben – zwei zentrale Risikofaktoren für die Motivation – untersucht. Diese werden nachfolgend näher beschrieben.

### 2.3 Prüfungsängstlichkeit – ein Risikofaktor für die Motivation

Auf der Basis eigener Erfahrungen und alltagspsychologischen Wissens ist nachvollziehbar, dass neben Erwartungen auch Emotionen einen entscheidenden Einfluss darauf haben, warum manche Handlungen aufgenommen und zu Ende gebracht, andere hingegen abgebrochen oder gar nicht erst begonnen werden (Pintrich, 2003). Im schulischen Kontext sind Emotionen wie Lernfreude, Hoffnung, Stolz, Ärger, Angst, Hoffnungslosigkeit, Scham und Langeweile allgegenwärtig und spielen eine zentrale Rolle für die Motivation, die Leistung und auch die persönliche Entwicklung der Schüler. Eine der am besten erforschten leistungsbezogenen Emotionen ist die Angst, speziell die sogenannte *Leistungs- beziehungsweise Prüfungsangst* („test anxiety“) (Pekrun, 2009).<sup>6</sup> Die Forschung reicht bis in die dreißiger Jahre des zwanzigsten Jahrhunderts zurück und umfasst inzwischen weit mehr als 1000 empirische Studien. Gründe für diese anhaltende Erforschung von Prüfungsangst sind das besonders häufige Vorkommen dieser Emotion und ihr Entwicklungsverlauf während der Schulzeit. Die Prävalenz von Prüfungsangst in der Schule variiert je nach Studie und Erhebungsinstrument zwischen 10 und 40 %. Mädchen, Minderheitengruppen und Schüler mit Lernschwierigkeiten sind besonders betroffen (Embse, Barterian & Segool, 2013). Studien belegen, dass die Prüfungsangst bei Eintritt in die Schule noch gering ausfällt, jedoch mit zunehmendem Alter und höheren Klassenstufen kontinuierlich zunimmt. Wäh-

---

<sup>6</sup> In der deutschen Literatur werden die Begriffe Leistungs- und Prüfungsangst oft synonym verwendet (z. B. Rost & Schermer, 2010). Nach Frenzel, Götz und Pekrun (2008) sind „Prüfungen typische, wenn auch nicht die einzigen Situationen, in denen Leistungsanforderungen an Personen gestellt werden. Prüfungsangst kann also als eine spezielle Form der Leistungsangst bewertet werden“ (S. 277). In der vorliegenden Arbeit wird die Prüfungsangst, speziell die Prüfungsängstlichkeit, als die spezifische Angst vor Prüfungen und Tests, untersucht. Eine Erläuterung der unterschiedlichen Definition von Prüfungsangst und -ängstlichkeit erfolgt im nachfolgenden Unterkapitel 2.3.1.

rend der Mittel- und Oberstufe stagniert die Prüfungsangst als ein Risikofaktor für die Motivation bei relativ hohen Werten (Pekrun, 2009).

Nachfolgend sollen die Begriffe Prüfungsangst und -ängstlichkeit zunächst definiert werden. Im Anschluss werden die Entstehungsbedingungen, verwandte Konstrukte und ausgewählte Erhebungsmethoden von Prüfungsangst und -ängstlichkeit dargestellt und schließlich Auswirkungen und Korrelate sowie mögliche Implikationen für die schulische Praxis beschrieben. Zuletzt wird eine spezifische schulische Intervention zur Reduktion der Prüfungsängstlichkeit, das sogenannte Transparenzpapier, gesondert vorgestellt, da genau diese Maßnahme auch Teil des FOSIS2-Programms ist.

### **2.3.1 Definition und theoretische Konzeption von Prüfungsangst und -ängstlichkeit**

„Bei Prüfungsangst handelt es sich um eine Angst, die sich auf Prüfungen bezieht, also auf Situationen einer zielgerichteten Erhebung und Bewertung von Leistungen, und damit um eine spezielle, auf eine bestimmte Situationsklasse gerichtete Form einer *Bewertungsangst*“ (Pekrun & Götz, 2006, S. 248). Bezüglich der Beschaffenheit von Prüfungsangst herrscht heute weitgehend darüber Einigkeit, dass sie aus verschiedenen Komponenten besteht (Pekrun, 2009; Wild et al., 2006): einer *affektiven* Komponente, die emotionale Angstanteile beschreibt (z. B. Nervosität, Unlust und Unsicherheit); einer *kognitiven* Komponente, die mit der jeweiligen Situation verbundene Gedanken beinhaltet (im speziellen Fall von Prüfungsangst beispielsweise das „Sichsorgen“ und „Grübeln“ über einen potentiellen Misserfolg und die damit verbundenen möglichen negativen Konsequenzen); einer *physiologischen* Komponente, die körperliche Reaktionen auslöst (z. B. Schwitzen, erhöhter Puls, Zittern und Übelkeit); einer *motivationalen* Komponente, die ein bestimmtes Verhalten nach sich zieht (z. B. Flucht und Vermeidung); und schließlich einer *expressiven* Komponente, die eine sorgenvolle Mimik, gebeugte Körperhaltung und beispielsweise ängstliche Ausrufe umfasst (Pekrun, 2009). Oft werden die affektiven und physiologischen Merkmale von Angst zu einer Komponente, der *Aufgeregtheit* („*emotionality*“), zusammengefasst und die mit der Prüfungsangst einhergehenden kognitiven und motivationalen Prozesse auch als *Besorgtheit* („*worry*“) bezeichnet (Frenzel, Götz, Pekrun & Fend, 2008). Nach Schwarzer (2000) ist die angesichts von Leistungsanforderungen auftretende Prüfungsangst vor allem durch diese beiden Merkmale Aufgeregtheit und Besorgtheit gekennzeichnet, wobei Letztere die für das Angsterleben entscheidendere Komponente darstellt.



Grundsätzlich wird zwischen der aktuellen *Prüfungsangst* (*state test anxiety*) in einer spezifischen Test- und Leistungssituation und der habituellen, dispositionellen *Prüfungsängstlichkeit* (*trait test anxiety*) im Sinne einer Persönlichkeitseigenschaft unterschieden. Letztere beschreibt die überdauernde Tendenz einer Person, in Prüfungssituationen generell eher ängstlich zu reagieren (Jerusalem, 1990; Pekrun, 2009; Schwarzer, 2000; Wild et al., 2006). In der pädagogisch-psychologischen Forschung wird häufiger die Prüfungsängstlichkeit als die Prüfungsangst untersucht (Wild et al., 2006). Auch im FOSIS2-Projekt wurde die Prüfungsängstlichkeit der Schüler und nicht ihre aktuelle Angst vor oder nach einem spezifischen Test erhoben. Bei Prüfungsängstlichen ist die kognitive Besorgtheits-Komponente besonders ausgeprägt, da bei ihnen von einer gewissen „kognitiven Voreingenommenheit gegenüber Anforderungssituationen, denen man sich nicht oder kaum gewachsen fühlt“ (Jerusalem, 1990, S. 42) ausgegangen werden muss. Deshalb wird im FOSIS2-Projekt als Maß für die Prüfungsängstlichkeit die Besorgtheit in Anlehnung an das „Test Anxiety Inventory“ (TAI) von Hodapp, Laux und Spielberger (1982) erfasst (eine genaue Beschreibung des Instruments erfolgt in Unterkapitel 6.4.1).

### 2.3.2 Entstehungsbedingungen von Prüfungsangst und -ängstlichkeit

An der Entstehung von Prüfungsangst und -ängstlichkeit sind sowohl *personale* Merkmale des Lerner selbst als auch Bedingungen der *sozialen* und der *Lernumwelt* beteiligt, die in Wechselwirkung miteinander treten und sich gegenseitig beeinflussen können (Pekrun & Götz, 2006).

#### *Merkmale der Person*

Im Hinblick auf die in der Person selbst zu lokalisierenden Ursachen für Emotionen wurde seit den sechziger Jahren das vorherrschende psychoanalytische Verständnis von Emotionen, deren Entstehen bis dahin mit intrapsychischen, unbewussten Konflikten erklärt wurde, durch kognitive Theorieansätze abgelöst (Pekrun, 2009): Die *Attributionstheorie* von Weiner (1985) und das *transaktionale Stressmodell* von Lazarus und Folkman (1984) sind Beispiele für solche kognitiven Erklärungsansätze von Emotionen. In Anlehnung an diese beiden Theorien und in Anknüpfung an die Erwartung-mal-Wert-Modelle hat Pekrun (1992) das sogenannte *Kontroll-Wert-Modell* zur Erklärung der Prüfungsangstentstehung entwickelt (Pekrun, 2009). Hiernach hängt die Ausbildung von Prüfungsangst maßgeblich ab von der subjektiv empfundenen *Kontrolle* in einer Prüfungssituation und von dem subjektiven *Wert*, also der individuell wahrgenommenen Bedeutung, die dem Ausgang der

Prüfung zugeschrieben wird. Wenn die eigenen Bewältigungsmöglichkeiten als unzureichend eingeschätzt werden, entsteht ein subjektiv erlebter Kontrollmangel und die Misserfolgserwartung fällt hoch aus. Wird einem solchen drohenden Misserfolg zusätzlich zu dem empfundenen Kontrollmangel noch ein hoher Wert, also eine große Bedeutung zugeschrieben, z. B. aufgrund damit verbundener antizipierter negativer Konsequenzen, entsteht eine aktuelle Prüfungsangst (z. B. Frenzel et al., 2008; Pekrun, 2009; Pekrun & Götz, 2006). Ist ein Schüler sich hingegen sicher, dass er bei einer Aufgabenbewältigung unter keinen Umständen scheitern wird oder ist das Ergebnis einer Prüfung für ihn völlig unwichtig, so wird dieser Schüler auch keine Angst empfinden (Pekrun, 1992). Umgekehrt wird ein Schüler auch keine Angst mehr zeigen, sondern mit Hoffnungslosigkeit reagieren, wenn das Eintreffen eines Misserfolgs absolut sicher ist und die Prüfung zudem eine besondere Wichtigkeit oder Bedeutung (Wert) aufweist. Die Prüfungsangst ist somit eine kurvilineare Funktion der Erfolgs- beziehungsweise Misserfolgswahrscheinlichkeit (Pekrun, 2009).

Neben der subjektiven Kontrolle und dem subjektiven Wert eines Prüfungsergebnisses, die beide im Lerner selbst angesiedelt sind, können auch Merkmale der *Sozial- und der Lernumwelt* zur Entstehung von Prüfungsangst beitragen (Embse et al., 2013), wobei vor allem solche Merkmale einen Einfluss haben, die sich wiederum auf die Kontroll- und Werteinschätzungen des Schülers auswirken (Pekrun & Götz, 2006).

### ***Merkmale der sozialen und der Lernumwelt***

Soziale Bedingungen, die die Entwicklung einer dispositionellen Prüfungsängstlichkeit begünstigen, sind beispielweise folgende: emotionale Kälte, besondere Strenge oder auch Desinteresse der Eltern (Rost & Schermer, 2010); insgesamt ein autoritärer Erziehungsstil oder auch eine inkonsequente Einhaltung von Regeln; zu hohe Leistungsanforderungen und damit verbundene Überforderung (Wild et al., 2006) sowie allgemein ein hoher familiärer und schulischer Leistungsdruck (Pekrun, 2009); die Bestrafung von Misserfolgen seitens Eltern und Lehrern, insbesondere Spott, Demütigung, Strafarbeiten oder auch der Entzug von Belohnungen und Zuwendung; das Vorherrschen einer sozialen Bezugsnorm, kompetitive Zielstrukturen und ein schlechtes Klassenklima (Rost & Schermer, 2010). Zudem zeigen sich Zusammenhänge zwischen der Prüfungsängstlichkeit des einzelnen Schülers und der durchschnittlichen Fähigkeit seiner Klasse: In einer leistungsstarken Klasse ist die durchschnittliche Prüfungsangst eines Schülers höher als in einer leistungsschwachen Klasse (Pekrun, 2009).

Seitens der Lernumwelt und speziell der Gestaltung von Prüfungen scheinen folgende Merkmale die Wahrscheinlichkeit der Ausbildung von Prüfungsangst zu erhöhen: Intransparenz von Anforderungen, Vorbereitungsmöglichkeiten und Bewertungskriterien, zu schwierige Aufgaben, eine mangelnde Struktur und Klarheit der Erklärungen des Lehrers (Pekrun, 2009), Zeitdruck, keine individualisierten Rückmeldungen, strikte Bewertungen, Selektion (Rost & Schermer, 2010) und einflussreiche, also bedeutsame Prüfungen (Pekrun & Götz, 2006). Auch das Antwortformat von Testfragen kann die Prüfungsangst beeinflussen: Multiple-Choice-Aufgaben scheinen weniger Angst auslösend zu sein als offene Fragen (Pekrun, 2009).<sup>7</sup>

### 2.3.3 Auswirkungen und Korrelate von Prüfungsangst- und -ängstlichkeit

Im Hinblick auf *Motivation*, *Lernen* und *Leistung* finden sich vor allem negative Effekte von Prüfungsangst, die sich – analog zu den Entstehungsbedingungen – wechselseitig beeinflussen können (Pekrun & Götz, 2006).

#### *Motivationale Auswirkungen und Korrelate*

Beginnend bei der für die Motivation bedeutsamen Ressource, der Selbstwirksamkeitserwartung, zeigt sich, dass Selbstwirksamkeitserwartung und Angst negativ miteinander korrelieren (Schwarzer, 2000; Schwarzer, 2004; Usher & Pajares, 2008). So erleben Schüler mit einer hohen Selbstwirksamkeitserwartung mehr positive Emotionen wie z. B. Zufriedenheit und Stolz. Hingegen erfahren Schüler mit gering ausgeprägten Kompetenzüberzeugungen in der Schule eher negative Emotionen wie z. B. Angst, Frustration und Ärger. Bandura (Bandura, 1988) geht davon aus, dass vor allem die Selbstwirksamkeitserwartung als erklärende Größe die Angst beeinflusst. Inzwischen gilt jedoch als gesichert, dass umgekehrt auch Angst und Ängstlichkeit mitbestimmen, wie kompetent sich Schüler einschätzen: Denn die Selbstwirksamkeitserwartung beeinflusst auch die Interpretation der eigenen emotionalen und körperlichen Erregung. Wenn das Gefühl der Angst, speziell die körperlichen Angstreaktionen (Schwarzer, 2000), als Beleg für einen eigenen Mangel an Fähigkeiten gedeutet wird, verringert dies auch die Selbstwirksamkeitserwartung (Schunk, 2003; Usher & Pajares, 2008).

---

<sup>7</sup> Angesichts der zahlreichen empirischen Studien zu Entstehungsbedingungen von Prüfungsangst und -ängstlichkeit haben die hier aufgezählten Merkmale nicht den Anspruch auf Vollständigkeit, sondern sollen vielmehr einen ersten Eindruck darüber vermitteln, dass vielfältige Faktoren zur Entstehung von Prüfungsangst beitragen können.

Neben dem negativen Zusammenhang zwischen Selbstwirksamkeitserwartung und Prüfungsangst belegen Studien bei Prüfungsängstlichen darüber hinaus auch eine ungünstige Entwicklung von anderen motivational relevanten Ressourcen wie beispielsweise dem Selbstkonzept der Begabung und Kontrollüberzeugungen (Pekrun, 2009). Angsterleben und Ängstlichkeit beeinträchtigen außerdem die Lernfreude, reduzieren die intrinsische Motivation (Wild et al., 2006) und begünstigen die Ausbildung anderer, die Motivation hemmender Risikofaktoren wie die Leistungsvermeidungszielorientierung (Pekrun, 2009). Auch die Annahmen des Kontroll-Wert-Modells von Pekrun (Pekrun, 1992) konnten weitestgehend empirisch bestätigt werden: Studien belegen hohe Korrelationen von Prüfungsangst, Misserfolgserwartung und der subjektiv empfundenen Bedeutung eines möglichen Misserfolgs. Zudem zeigen sich ebenfalls Zusammenhänge zwischen Prüfungsangst und Kausalattributionen (Wild et al., 2006): Prüfungsängstliche attribuieren Misserfolge eher ungünstig internal stabil auf einen eigenen Mangel an Fähigkeiten und neigen bei Misserfolgen zu Selbstbeschuldigungen und abwertenden Gedanken (Schwarzer, 2000). Dieser Attributionsstil verringert wiederum die Erfolgserwartung im Hinblick auf die Bewältigung zukünftiger Anforderungen und kann somit die Prüfungsangst noch verstärken (Wild et al., 2006).

### ***Auswirkungen auf Lernen und Leistung***

Hinsichtlich der Anwendung von Lernstrategien konnte empirisch nachgewiesen werden, dass Angst vor allem mit dem Einsatz von Oberflächenstrategien, wie z. B. Wiederholung, einhergeht. Lernfreude korreliert hingegen mit dem Einsatz von kognitiv flexibleren Tiefen-Strategien wie Elaboration (Frenzel et al., 2008; Pekrun & Götz, 2006; Wild et al., 2006). Auch die Fähigkeit zum selbstregulierten Lernen verringert sich durch Prüfungsangst, und Prüfungsängstliche neigen eher dazu, bei der Zielsetzung, Steuerung und Bewertung des eigenen Lernens Anweisungen von Autoritätspersonen Folge zu leisten (Pekrun & Götz, 2006). Hochängstliche fallen darüber hinaus oft durch eine mangelnde Aufgabenzuwendung und unstrukturierte Aufgabenbearbeitung auf (Rost & Schermer, 2010). In Abhängigkeit von der Aufgabenschwierigkeit besteht auch zwischen Prüfungsangst und Leistung ein bedeutsamer negativer Zusammenhang: Bei komplexen und schwierigen Aufgaben verschlechtern sich in der Regel die Leistungen durch das Auftreten von Prüfungsangst (z. B. Hembree, 1988; Zeidner, 1998, 2007). Im Gegensatz dazu zeigt sich bei einfachen Wiederholungsaufgaben nicht notwendigerweise ein negativer Einfluss von Prüfungsangst auf die Leistung (Pekrun, 2009).

Diese Effekte können theoretisch mittels der erstmals von Wine (1971) beschriebenen *Aufmerksamkeits-Defizit-Hypothese* erklärt werden (Pekrun, 2009): Prüfungsangst absorbiert durch die kognitive Angstkomponente, durch das ständige Sichsorgen und Grübeln über einen möglichen Misserfolg sowie durch aufgaben-irrelevante Gedanken ein hohes Maß an Aufmerksamkeit. Entsprechend verfügen prüfungsängstliche Schüler insbesondere bei schwierigen Aufgaben, die das Arbeitsgedächtnis in Anspruch nehmen, über eine geringere kognitive Kapazität für die eigentliche Problem- und Aufgabenbewältigung sowie über eine schlechtere Informationsverarbeitung als Schüler, die einer Prüfungssituation ohne Angst begegnen. Dadurch wiederum reduziert sich für Prüfungsängstliche die Erfolgswahrscheinlichkeit, wodurch sowohl die Misserfolgserwartung als auch die Angst gesteigert werden. Es besteht die Gefahr, in einen Teufelskreis aus Angst, hoher Misserfolgserwartung, realer Erfahrung von Misserfolgen beziehungsweise mangelnden Leistungen zu geraten und somit auch zukünftigen Prüfungssituationen mit noch mehr Angst zu begegnen (Pekrun, 2009).

Ein verwandter Teufelskreis, der ebenfalls die Auswirkungen von Prüfungsangst auf Lernen und Leistung beschreibt, wird im Rahmen der *Begabungs-Defizit-Hypothese* angenommen. Vertreter dieser Hypothese halten einen tatsächlichen Mangel an Fähigkeiten und Begabungen für den Ausgangspunkt der Entstehung von Prüfungsangst (Zeidner, 1998): Schüler mit mangelnden Fähigkeiten erleben bei schwierigen, komplexen Aufgaben auch mit einer höheren Wahrscheinlichkeit Misserfolge, antizipieren weitere Misserfolge für die Zukunft und reagieren angesichts von schwierigen Aufgaben deshalb auch eher mit Prüfungsangst als fähige Schüler. Empirisch konnten beide Annahmen bestätigt werden. Prüfungsangst geht mit aufgaben-irrelevantem Denken einher und lenkt dadurch von der eigentlichen Problemlösung bei komplexen Aufgaben ab. Zudem zeigen Schüler mit eher geringen Fähigkeiten häufiger Prüfungsangst als Schüler mit einer hohen Ausprägung verschiedener Fähigkeiten (Pekrun, 2009). Da Prüfungsängstliche auf Anforderungen vermehrt mit Hilflosigkeit und Unsicherheit reagieren und öfter krank sind, beziehungsweise höhere Fehlzeiten aufweisen als Niedrigängstliche, werden Prüfungsängstliche von ihrem sozialen Umfeld in ihren kognitiven Fähigkeiten zudem häufig unterschätzt und als „unangepasst“ wahrgenommen (Rost & Schermer, 2010).

### ***Gibt es positive Auswirkungen von Prüfungsangst und -ängstlichkeit?***

Neben den dominierenden negativen Effekten kann Prüfungsangst auch positive Konsequenzen nach sich ziehen. So ist es unter Umständen möglich, dass sich durch die mit der Angst einhergehende höhere Misserfolgsvermeidungsmotivation die Anstrengungsbereitschaft verstärkt (Frenzel et al., 2008). Insbesondere in Situationen, wo eine Vermeidung der Prüfungssituation den drohenden Misserfolg nicht verhindern kann (z. B. Durchfallen durch die Prüfung bei unbegründetem Nichterscheinen), kann Angst auch die Motivation zu Lernen verbessern und die Anstrengungsbereitschaft und Ausdauer steigern (Wild et al., 2006). Insgesamt überwiegen bei der Mehrheit der Schüler jedoch die negativen Effekte von Prüfungsangst auf die Motivation und damit auch auf das Lernverhalten und die Leistung. Zudem kann insbesondere die dispositionelle Prüfungsängstlichkeit auch das psychische Wohlbefinden, die körperliche Gesundheit sowie die soziale Eingebundenheit deutlich beeinträchtigen (Pekrun, 2009).

#### **2.3.4 Mit Prüfungsangst und -ängstlichkeit verwandte Konstrukte**

Ebenfalls typisch für den Kontext Schule und verwandt mit der Prüfungsangst sind die *soziale Angst* beziehungsweise *Ängstlichkeit* und die *Schulphobie*. Die soziale Angst entsteht durch die erlebte Bedrohung des eigenen Selbstwerts. Sie kann sich in Scham, Schüchternheit, Verlegenheit und Publikumsangst äußern (Jerusalem, 1990; Schwarzer, 2000). In allen Fällen entsteht die Angst durch die antizipierte Abwertung oder negative Behandlung in sozialen Situationen (Jerusalem, 1990). Auch bei der Prüfungsangst ist die Bedrohung des eigenen Selbstwerts eine zentrale Erlebniskomponente. Allerdings basiert diese wahrgenommene Bedrohung auf der Furcht vor einem Misserfolg in Leistungssituationen, deren erfolgreiche Bewältigung einen hohen Stellenwert für den eigenen Selbstwert hat. Da viele Prüfungen auch eine soziale Bewertung mit sich bringen, können soziale und Prüfungsangst gemeinsam beziehungsweise vermischt auftreten (Jerusalem, 1990; Schwarzer, 2000).<sup>8</sup> Prüfungsängstliche neigen sogar dazu, in Leistungssituationen auf soziale Bewertungsaspekte zu achten und diese zu befürchten. Somit sind prüfungsängstliche Personen für soziale Ängste besonders empfänglich (Schwarzer, 2000). Nach Rost und Schermer (2010) muss außerdem die Schulphobie von

---

<sup>8</sup> Ein spezieller Fall der Prüfungsangst ist zudem die Mathematikangst, die eine Angst vor Zahlen und Formeln beschreibt. Wie auch die Selbstwirksamkeit hinsichtlich ihrer Spezifität und Generalität differenziert werden kann, so können auch verschiedenste situationsspezifische Ängste unterschieden werden (z. B. Angst vorm Fahrstuhlfahren, vorm Fliegen, vor Unfällen) und eine Ausdifferenzierung der Prüfungsangst nach weiteren Schulfächern ist denkbar (Schwarzer, 2000).

der Prüfungsangst unterschieden werden. Dabei handelt es sich um die „behandlungsbedürftige Form länger anhaltender Schulbesuchsverweigerung“ (Rost & Schermer, 2010, S. 451), die mit einer starken Furcht, mit Trennungsängsten und oft auch mit körperlichen Beschwerden einhergeht. Auch Angst- und Stresserleben weisen einige Überschneidungen auf. Stresserleben entsteht dann, wenn die Anforderungen einer Situation die subjektiv eingeschätzten Ressourcen zur Bewältigung dieser Situation übersteigen (Pekrun & Götz, 2006). Nach dem Kontroll-Wert-Modell von Pekrun (1992) entsteht Prüfungsangst ebenfalls durch einen Kontrollverlust aufgrund von als unzureichend eingeschätzten Fähigkeiten verbunden mit einem hohen Wert der Prüfung. Prüfungsangst kann somit als eine spezifische Form des Stresserlebens vor und während Prüfungen verstanden werden und überschneidet sich vor allem mit der wahrgenommenen Bedrohung, einer Komponente des Stresserlebens (der Zusammenhang zwischen dieser Komponente des Stresserlebens und der Prüfungsängstlichkeit wird im Abschnitt 2.4.5 ausführlicher erläutert).

### **2.3.5 Erfassung von Prüfungsangst und -ängstlichkeit**

Lange Zeit wurde Prüfungsangst als eindimensionales Konstrukt gemessen (z. B. „Test Anxiety Questionnaire – TAQ“ von Mandler und Sarason, 1952 oder „Test Anxiety Scale for Children – TASC“ von Sarason, Davidson, Lighthall und Waite, 1958). Der theoretischen Unterscheidung von kognitiven und emotionalen Angstanteilen entsprechend entwickelten Liebert und Morris (1967) den „Worry-Emotionality-Questionnaire“, der beide Aspekte der situativen Zustandsangst erfasst. International besonders häufig verwendet wurde das „Test Anxiety Inventory – TAI“ von Spielberger (1980), das die Komponenten Besorgtheit und Aufgeregtheit als Bestandteile der Ängstlichkeit im Sinne einer überdauernden Persönlichkeitseigenschaft misst (Schwarzer, 2000). Die in dieser Arbeit verwendete Skala zur Erfassung der Prüfungsängstlichkeit wurde in Anlehnung an die Besorgtheits-Komponente des deutschen TAI von Hodapp et al. (1982) entwickelt (siehe Abschnitt 6.4.1). Auch biometrische Methoden (z. B. fMRI, EEG) werden zur Erfassung der Prüfungsangst eingesetzt (Pekrun, 2009). Da diese mit einem vergleichsweise hohen Aufwand verbunden sind und vor allem die aktuelle Prüfungsangst in einer Situation messen, wurden im FOSIS2-Projekt keine biometrischen Methoden verwendet.

### **2.3.6 Implikationen für die Praxis: Reduktion von Prüfungsangst und -ängstlichkeit**

Die Entstehungsbedingungen und Auswirkungen von Prüfungsangst erlauben Schlussfolgerungen für die Entwicklung von Angst reduzierenden Interventionen. Diese können unterteilt werden in *personenbezogene* Maßnahmen, die beim Schüler selbst ansetzen (Verhaltensprävention) und Maßnahmen, die das Ziel verfolgen, die *soziale* und die *Lernumwelt* des Schülers zu verändern (Verhältnisprävention). Die Interventionen haben sich vor allem dann als Angst reduzierend erwiesen, wenn sie Unsicherheiten verringern und Erfolgserwartungen verstärken (Rost & Schermer, 2010).

#### ***Maßnahmen zur Veränderung der Person und des Verhaltens***

Bei der Prüfungsängstlichkeit handelt es sich nicht um eine klinische Störung, die unbedingt einen therapeutischen Behandlungsbedarf nach sich zieht. Sie kann jedoch eine so starke Ausprägung annehmen, dass sie therapeutisch behandelt werden muss (Schwarzer, 2000). Dafür haben sich aus verhaltenstherapeutischer Perspektive vor allem das Biofeedback, Entspannungstechniken (Frenzel et al., 2008) und bei einer schwer ausgeprägten Form der Prüfungsängstlichkeit auch die klassische Desensibilisierung als wirksame Interventionen erwiesen (Frenzel et al., 2008; Rost & Schermer, 2010). Die Desensibilisierung sollte jedoch nicht allein angewendet werden, sondern mit dem Training von Lern-, Arbeits- und Selbstregulationsstrategien sowie mit kognitiv-verhaltenstherapeutischen Maßnahmen zur Angstreduktion kombiniert werden (Rost & Schermer, 2010). Beispiele für Letztgenannte sind die kognitive Umstrukturierung und das Ersetzen angst-typischer katastrophisierender Gedanken durch hoffnungsvolle, zuversichtliche Erfolgserwartungen und aufgabenbezogene Kognitionen (Frenzel et al., 2008). Diese Empfehlungen bestätigen auch Embse et al. (2013) in ihrem systematischen Literaturüberblick zu Prüfungsangst reduzierenden Interventionen für Kinder und Jugendliche. Die Autoren fassen hier die Ergebnisse von zehn experimentellen oder quasi-experimentellen Studien zusammen und identifizieren sowohl rein *verhaltensbezogene* als auch *kognitive* Interventionen sowie eine *Kombination* beider Ansätze als geeignete Maßnahmen zur Reduktion der Prüfungsangst. Rein *verhaltensbezogene* Maßnahmen sind Entspannungsverfahren (Larson, El Ramahi, Conn, Estes & Ghibellini, 2010) und die systematische Desensibilisierung (Egbochuku & Obodo, 2005). Eine rein *kognitive* Maßnahme wurde in einer Studie umgesetzt, in der eine Imaginationsübung („*priming*“) mit Schülern evaluiert wurde. Vor einem Test wurden die Schüler aufgefordert, sich eine kompetente Person vorzustellen und deren Fähigkeiten zu



notieren, um so die eigene Kompetenzwahrnehmung zu stärken und die Angst zu reduzieren (Lang & Lang, 2010). Schließlich stellen Embse et al. (2013) auch noch eine Studie vor, in der eine Kombination kognitiv-verhaltenstherapeutischer Interventionen zum Einsatz kam, wie Methoden zur kognitiven Umstrukturierung negativer Gedanken verbunden mit Entspannungstechniken (Gregor, 2005). In weiteren in dem Review zitierten Studien finden sich ebenfalls Hinweise auf die Wirksamkeit von Biofeedbackmethoden (Bradley et al., 2010) und auf die Angst reduzierende Wirkung des Trainings von Lern- und Testbearbeitungsstrategien (Carter et al., 2005). Da die von Embse et al. (2013) zusammengefassten Studien jedoch unterschiedliche Teilziele verfolgen, von der Reduktion der aktuellen Zustands-Prüfungsangst vor einem Test bis hin zur Verringerung der Besorgtheits-Komponente der Prüfungsängstlichkeit, und da die Studien in unterschiedlichen Ländern, an verschiedenen Schularten und Altersgruppen durchgeführt wurden, sind die zitierten Interventionen weder vergleichbar noch auf die schulische Praxis generalisierbar. Embse et al. (2013) weisen deshalb darauf hin, dass weitere nationale Forschung nötig ist, die spezifischere Schlussfolgerungen für eine angstfreie schulische Praxis erlaubt.

### ***Maßnahmen zur Veränderung der sozialen und der Lernumwelt***

Hinsichtlich der sozialen und der Lernumwelt haben sich hohe Leistungserwartungen von bedeutsamen Personen (Eltern, Lehrer, Peers), negative Rückmeldungen und Bestrafungen im Falle eines Misserfolgs als Angst steigernd erwiesen (Pekrun, 2009). Entsprechend trägt eine intensive, offene und wertschätzende Kommunikation über schulische und Leistungssituationen zur Angstprävention bei. Das elterliche Verhalten im Hinblick auf Rückmeldungen zu Prüfungsleistungen sollte zudem möglichst konsistent und transparent sein und überfordernde, Druck erzeugende Ansprüche vermeiden. Demgegenüber hilft es nicht, auf jegliche Leistungsanforderungen zu verzichten. Vielmehr müssen die Leistungsansprüche angemessen sein, das heißt den Fähigkeiten und Ressourcen des Schülers entsprechen (Rost & Schermer, 2010). Auch autoritäre Erziehungs- und Unterrichtsstile der Lehrkraft erhöhen die Angst, wohingegen kreative, aufgabenorientierte Unterrichtsmethoden die Angst verringern. Zudem können die Emotionen der Lehrkraft selbst helfen, die Prüfungsangst zu reduzieren: Werden Schüler von Lehrern unterrichtet, die Freude am Unterrichten haben und Begeisterung zeigen, weisen auch die Schüler mehr Lernfreude und weniger Prüfungsangst auf. Generell wirkt sich ein kompetitives Klassenklima negativ auf die Prüfungsangstentwicklung aus (Pekrun, 2009) und ein positives, kooperatives Klassenklima kann der Angstentstehung vorbeugen (Wild et al.,

2006). Anders als zu erwarten wäre, bewirken ein gutes, kooperatives Klassenklima und auch die soziale Unterstützung durch Lehrer und Eltern jedoch nicht immer die gewünschten positiven Effekte im Sinne einer Angstreduktion. Dies mag an der oft von den Schülern als aufgezwungen erlebten Unterstützung liegen, die eigene Kontroll- und Autonomiebemühungen reduzieren und möglicherweise das gegenteilige Gefühl von Angst und Hilflosigkeit nach sich ziehen kann (Pekrun, 2009). In diesem Sinn erweist sich die Förderung von Selbstbestimmung im Unterricht, speziell die Möglichkeit der Wahl zwischen verschiedenen Testaufgaben sowie das Angebot, einen Test zu wiederholen, als Angst reduzierend. Denn durch Wahl- und Wiederholungsmöglichkeiten werden die wahrgenommene Kontrolle und auch die Erfolgswahrscheinlichkeit bei der Bewältigung von Lernaufgaben und Prüfungen deutlich gesteigert.

Auch die Transparenz von Anforderungen, Vorbereitungshilfen und Bewertungskriterien sowie eine klare Aufgabeninstruktion seitens der Lehrkraft und generell die Erhöhung von Erfolgserfahrungen reduzieren die Angst vor Prüfungen beziehungsweise beugen einer Angentstehung vor (Pekrun, 2009). Hinsichtlich der Förderung von Transparenz zur Reduktion der Prüfungsangst und langfristig auch der Prüfungsängstlichkeit soll im Folgenden eine spezifische Intervention beschrieben werden, die auch im FOSIS2-Projekt zum Einsatz kam: In mehreren Interventionsstudien wurde in der Arbeitsgruppe um Strittmatter zur Reduktion der Prüfungsangst der Einsatz eines sogenannten *Transparenzpapiers* vor Klassenarbeiten erprobt (z. B. Jacobs, 1987; Jacobs, 1997; Jacobs & Bedersdorfer, 1984; Jacobs, Bedersdorfer & Bohse-Wagner, 1983; Strittmatter, 1997). Nach Jacobs (1997) werden mit dem Begriff Angst umgangssprachlich auch Unsicherheit und Ungewissheit assoziiert. Speziell in Prüfungssituationen kann Ungewissheit darüber bestehen, was in welcher Weise abgeprüft wird, wie eine effektive Vorbereitung gestaltet werden kann und wie die Prüfung bewertet wird. Um diese Unklarheiten zu beseitigen, stattdessen Transparenz und Klarheit zu steigern und dadurch auch die Angst vor Prüfungen zu reduzieren, wurde das Transparenzpapier entwickelt. Dabei handelt es sich um ein Informationspapier, das in der Schule rechtzeitig vor Prüfungen ausgeteilt wird und über die folgenden Bereiche informieren sollte:

- *Anforderungen und Gestaltung der Leistungsüberprüfung*
- *Kriterien der Leistungsbewertung*
- *Konsequenzen der Leistungsbewertung*

In den Bereich der Anforderungen und Gestaltung der Leistungsüberprüfung fallen neben den Themen und Inhalten einer Prüfung auch die Prüfungsbedingungen und -modalitäten: Das Transparenzpapier sollte folglich auch Vorbereitungshilfen nennen und in einer Klausur verwendete Aufgabentypen bekannt geben. Zudem informiert ein Transparenzpapier über die Kriterien der Leistungsbewertung, die Gewichtung der einzelnen Aufgaben sowie die Zuordnung der Gesamtpunktzahlen zu Noten. Schließlich enthält ein vollständiges Transparenzpapier auch noch Hinweise auf die Konsequenzen der Leistungsbewertung. So sollte es beispielsweise den Einfluss der Klausurnote auf das Endergebnis (z. B. die Zeugnisnote) beschreiben (Strittmatter, 1997).

Die Konzeption des Transparenzpapiers gründet auf angsttheoretischen Annahmen, die dem Kontroll-Wert-Modell zur Angstentstehung von Pekrun (Pekrun, 1992) ähneln. Die Prüfungsangstentstehung beginnt nach Jacobs (1997) mit einem wahrgenommenen Kompetenzdefizit, das aus einer mangelnden Passung zwischen dem Anspruchsniveau einer Person und subjektiv eingeschätzten Kompetenzen resultiert. Dadurch steigt die Misserfolgsbefürchtung, die in Kombination mit antizipierten negativen Konsequenzen des drohenden Misserfolgs dann die aktuelle Angst vor und in der Prüfung auslöst. Entsprechend diesen Annahmen ist zu erwarten, dass sich Prüfungsangst durch die Senkung des Anspruchsniveaus, durch die Steigerung des Kompetenzerlebens und die Reduktion negativ antizipierter Konsequenzen verringern lässt. Der von der Arbeitsgruppe um Strittmatter theoretisch postulierte Wirkungsweg von einer gesteigerten Transparenz zu einer verringerten Prüfungsangst beginnt mit einer „objektiven“ Erhöhung der Transparenz: Wird seitens des Lehrers eine Prüfungssituation objektiv transparenter gestaltet, sollte dies bei den Schülern auch das „subjektive“ Transparenzerleben erhöhen. Jacobs (1997) nimmt des Weiteren an, dass diese subjektive Transparenz dazu beiträgt, dass die Schüler ihre Kompetenzen zur Bewältigung der Prüfung positiver einschätzen und sich somit das wahrgenommene Kompetenzdefizit sowie die daraus resultierende Misserfolgsbefürchtung und schließlich auch die aktuelle Prüfungsangst verringern. Inwieweit die theoretischen Annahmen zutreffen, wurde in verschiedenen empirischen Studien überprüft: Jacobs und Bendersdorfer (1984) erprobten den Einsatz des Transparenzpapiers an Lehramtsstudierenden, wobei sie einer Gruppe von Studierenden zwei Wochen vor der Klausur ein Transparenzpapier austeilten (Experimentalgruppe). Studierende einer Kontrollgruppe erhielten kein Treatment. Mittels eines Prä- und Post-Designs wurde die Wirksamkeit des Transparenzpapiers evaluiert. Die Ergebnisse der Studie belegen einen eindeutigen Anstieg der subjektiv wahrgenommenen Transparenz und liefern erste Hinweise auf eine Angst reduzierende

Wirkung des Transparenzpapiers. In einer anderen Studie wurde der Einsatz des Transparenzpapiers in der Schule an Schülern der 12. Klasse untersucht (Jacobs et al., 1983). Neben der Experimental- und der reinen Kontrollgruppe ohne Treatment wurde zusätzlich eine Placebogruppe eingeführt. Die Schüler dieser Gruppe erhielten vor der Klassenarbeit ein Informationspapier mit unspezifischen Hinweisen über Prüfungen im Allgemeinen (z. B. Erläuterungen zum pädagogischen Anspruch von Prüfungen). Im Vergleich zur Kontrollgruppe zeigen sich bei den Schülern der Experimentalgruppe ein signifikanter Anstieg der subjektiv erlebten Transparenz, eine Verbesserung der Leistungen und eine bedeutsame Reduktion der Misserfolgsbefürchtung und der Prüfungsangst. Die beiden letzten Effekte können im Experimental- und Placebogruppenvergleich jedoch nicht bestätigt werden. Auch im Saarbrücker Schulangstprojekt (Strittmatter, 1997) wurden Transparenzpapiere mit dem Ziel der Angstreduktion verwendet. Jacobs (1987), der den Einsatz der Transparenzpapiere in diesem Projekt evaluiert, belegt wiederum eine bedeutsame Steigerung der objektiven und subjektiven Transparenz, kann jedoch keinen signifikanten Anstieg des subjektiven Kompetenzerlebens oder eine Reduktion der Angst nachweisen. Der theoretisch angenommene Wirkungsweg von der Transparenzsteigerung zur Angstreduktion kann somit nur teilweise bestätigt werden. Jacobs (1997) begründet dies damit, dass ein Transparenzpapier nur dann die angenommenen Wirkungen erzielt, wenn es auch wirklich effektive Vorbereitungshilfen enthält und tatsächlich die Wahrscheinlichkeit erhöht, die Prüfungsanforderungen zu erfüllen. Dies ist vor allem dann gegeben, wenn Lehrziele abgefragt werden, die wenig Vorwissen voraussetzen. Ist für die erfolgreiche Bewältigung einer Prüfung neben dem tatsächlich abgeprüften Wissen ein umfangreiches Vorwissen erforderlich, schwächt dies vermutlich die Wirkung des Transparenzpapiers.

Stresserleben und Prüfungsangst werden als eng verwandte Konstrukte verstanden (Pekrun & Götz, 2006). In diesem Sinne lässt sich vermuten, dass der Einsatz von Transparenzpapieren nicht nur zu einer Reduktion des Angst-, sondern auch des Stresserlebens beiträgt. Zudem sollen Transparenzpapiere das Kompetenzerleben steigern und Schüler darin unterstützen, eine komplexe und schwierige Aufgabe, wie die Vorbereitung auf eine Klassenarbeit, in Nahziele zu untergliedern und Schritt für Schritt zu bewältigen. Die Untergliederung einer Aufgabe in Teilziele erhöht die Wahrscheinlichkeit für die Erfahrung eigener Erfolge (Schunk, 1995a). Somit sollte sich der regelmäßige Einsatz von Transparenzpapieren im Unterricht auch förderlich auf die schulische Selbstwirksamkeitserwartung von Schülern auswirken. Diese Annahme über die positiven Auswirkungen des Einsatzes von

Transparenzpapieren auf die Kompetenzüberzeugungen konnte empirisch bereits belegt werden (Spinath, 2008).

### **2.3.7 Zusammenfassung: Prüfungsangst und -ängstlichkeit**

In der vorliegenden Untersuchung wird neben der schulischen Selbstwirksamkeitserwartung auch die Prüfungsängstlichkeit als Kriterium zur Evaluation des FOSIS2-Programms herangezogen. Genau wie die aktuell in einer spezifischen Testsituation auftretende Prüfungsangst stellt auch die dispositionelle Prüfungsängstlichkeit (Pekrun, 2009) einen Risikofaktor für die schulische Motivation dar, der sich beispielsweise negativ auf motivationale Ressourcen wie die Selbstwirksamkeitserwartung (z. B. Usher & Pajares, 2008) und Lernfähigkeiten (z. B. die Anwendung von Tiefenstrategien oder die Fähigkeit zum selbst-regulierten Lernen; Frenzel et al., 2008) auswirkt. Zur Entstehung von Prüfungsangst und -ängstlichkeit tragen sowohl Merkmale der Person bei (siehe Kontroll-Wert-Modell der Prüfungsangst, Pekrun, 1992) als auch Merkmale der sozialen und der Lernumwelt wie ein autoritärer Erziehungsstil (Wild et al., 2006), Leistungsdruck, Intransparenz (Pekrun, 2009) und die Bestrafung von Misserfolgen (Rost, 2010). Die FOSIS2-Maßnahmen verfolgen das Ziel, einige der möglichen Entstehungsbedingungen von Prüfungsangst und -ängstlichkeit positiv zu beeinflussen. So soll beispielsweise der Einsatz von Transparenzpapieren vor Klassenarbeiten, durch die rechtzeitige, schriftliche Informierung über Anforderungen, Vorbereitungsmöglichkeiten, Kriterien der Leistungsbewertung und Konsequenzen die Transparenz fördern (z. B. Jacobs, 1997; Strittmatter, 1997,).

Neben der Prüfungsängstlichkeit wird in dieser Arbeit auch das Stresserleben mit seinen Unterkomponenten *wahrgenommene Bedrohung* und *wahrgenommener Verlust* als weitere Risikofaktoren sowie der Komponente *wahrgenommene Herausforderung* als eine weitere Ressource für die Motivation untersucht.

## 2.4 Schulbezogenes Stresserleben – Herausforderungserleben als Resource, Bedrohungs- und Verlusterleben als Risikofaktoren für die Motivation

Immer wieder wird in den Medien ein zunehmendes *Stresserleben* im Kindes- und Jugendalter<sup>9</sup> konstatiert. Dies gilt insbesondere für den Bereich Schule. So erschien beispielsweise am 22. April 2013 *Der Spiegel* mit dem Titel „Generation Stress – Wenn Schule krank macht“ (Augstein, 2013). Auch das ZDF widmete sich dem Thema und zeigte Dokumentationen zu den Auswirkungen von schulischem Stresserleben wie die Sendung mit dem Titel „Schüler in der Leistungsfalle“ (Wegner, 2014) oder die Reportage „Stressfaktor Schule“ (Abel, 2013). Solche und andere Berichte der Medien heben das Stresserleben als einen zentralen Risikofaktor für Motivation, Lernen und Leistung hervor, der immer mehr Kinder und Jugendliche zu betreffen scheint.

Hat das Stresserleben im Kindes- und Jugendalter im Laufe der letzten Jahrzehnte wirklich zugenommen? Die Betrachtung der Veränderung verschiedener soziokultureller Faktoren, die für das Kindes- und Jugendalter relevant sind, liefert Hinweise auf eine Zunahme des Stresserlebens bei Kindern und Jugendlichen (Lohaus, Domsch & Fridrici, 2007; Seiffge-Krenke et al., 2012). So hat sich die Vielfalt an möglichen Freizeitaktivitäten in den letzten Jahrzehnten enorm vergrößert. Viele Kinder besitzen schon im Grundschulalter einen vollen Terminplan und verfolgen ein immer größer werdendes Spektrum verschiedener Hobbys und Aktivitäten (z. B. das Erlernen von Musikinstrumenten und Trainieren verschiedener Sportarten). Dadurch und auch aufgrund der gleichzeitigen Zunahme an medialen Freizeitangeboten durch Computer und Fernsehen haben das ungeplante und spontane Spielen im Vergleich zu früher abgenommen und mit Zeitnot und dem Mangel an kreativen Freiräumen verbundene Belastungen zugenommen. Sowohl in den Freizeitbereichen als auch in Schule und Familie ist zudem ein allgemeiner Trend hin zur Individualisierung zu beobachten. Dahinter verbirgt sich das Ziel, die Begabungen und Kompetenzen des einzelnen Kindes in allen Bereichen so zu fördern, dass sie sich optimal entfalten können. Gerade dieser Wunsch nach dem Optimum trägt jedoch statt zur freien Entfaltung nicht selten zu einer wachsenden Konkurrenz in Schule und Freizeit bei und geht mit hohen Erwartungen

---

<sup>9</sup> Zielgruppe der FOSIS2-Interventionen waren Jugendliche im durchschnittlichen Alter von 12 bis 14 Jahren. Entwicklungspsychologisch fällt diese Altersspanne in das Jugendalter. „Als Jugend im Sinne der Entwicklungspsychologie bezeichnet man die Zeit zwischen der Pubertät und dem Ende des zweiten Lebensjahrzehnts (circa 10. bis 20. Lebensjahr)“ (Silbereisen, 2012, S. 236). Bei der Darstellung des theoretischen Hintergrunds zum Stresserleben werden – wenn möglich – die für das Jugendalter typischen Aspekte und Prozesse erläutert.

seitens der Lehrer, Mitschüler und Eltern sowie einem damit verbundenem hohen Leistungsdruck einher (Lohaus et al., 2007). Tatsächlich belegen zahlreiche Studien die wachsenden Anforderungen insbesondere im Kontext Schule. Konkurrenz und Leistungsdruck zählen somit zu den häufigsten schulbezogenen Stressauslösern (*Stressoren*) (Seiffge-Krenke, 2008; Seiffge-Krenke et al., 2012).

Durch veränderte Familienstrukturen, wie z. B. dem Trend zur Ein-Kind- oder Zwei-Kind-Familie, verteilen sich zudem die familiären Erwartungen nicht mehr auf mehrere Kinder; dies erhöht zusätzlich den Druck auf das einzelne Kind. Hinzu kommen Verschiebungen in den Selektions- und Allokationsmechanismen des Schulwesens sowie in Bezug auf die Bildungsaspiration: So gewinnt beispielsweise das Abitur im Hinblick auf die Ausbildungs- und Arbeitsplatzsicherheit an Relevanz, wohingegen Real- und Hauptschulabschlüsse eine wachsende Unsicherheit bezogen auf den weiteren Bildungsweg erzeugen (Lohaus et al., 2007) und Zukunftsängste im Jugendalter steigern können (Seiffge-Krenke, 2008; Seiffge-Krenke et al., 2012). Natürlich treffen nicht all die geschilderten Belastungen auf alle Kinder und Jugendlichen zu – insgesamt kann aber von einem wachsenden Belastungs- und Stresserleben im Kindes- und Jugendalter ausgegangen werden (Lohaus et al., 2007). Diese Vermutung schlägt sich auch in der häufigen Verwendung des Begriffs *Stress* in der Alltags- und Umgangssprache nieder (Klein-Heßling & Lohaus, 2012). So „stresst“ beispielsweise ein Lehrer oder es wird über die „stressigen“ Eltern geklagt (Lohaus et al., 2007, S.4). Nach Klein-Heßling und Lohaus (2012) geben von 342 befragten Kindern und Jugendlichen 72 % der Sieben- bis Elfjährigen und 81 % der Zwölf- bis Achtzehnjährigen an, regelmäßig Stress zu erleben. Die von den Kindern und Jugendlichen benannten Stressoren entstammen vornehmlich aus Leistungssituationen im schulischen Kontext (z. B. bevorstehende Klassenarbeiten oder zu viele, überfordernde Hausaufgaben) und aus sozialen Situationen (z. B. Konflikte mit Eltern und Freunden) (Klein-Heßling & Lohaus, 2012; Lohaus, 2006; Lohaus et al., 2007).

Nachfolgend sollen der Begriff Stress zunächst definiert und theoretische Konzeptionen von Stress und Stresserleben erläutert werden. Im Anschluss werden die Entstehungsbedingungen, die Auswirkungen und Korrelate von Stress sowie Ressourcen und Möglichkeiten zur Stressbewältigung beschrieben. Darauf folgt eine Darstellung des Zusammenhangs von Stress und Angst als verwandte Konstrukte und eine Schilderung ausgewählter Erhebungsmethoden von Stress im Kindes- und Jugendalter. Das Kapitel schließt mit einer Ausführung möglicher Interventionen zur Stressprävention für die schulische Praxis.

### 2.4.1 Definition und theoretische Konzeption von Stresserleben

In der Stressforschung existiert kein einheitliches Grundverständnis von Stress und Stresserleben (Jerusalem, 1990). Ähnlich wie bei dem Begriff Motivation handelt es sich bei dem Begriff Stress also auch um ein abstraktes Konstrukt, unter dem verschiedene Erlebens- und Verhaltensaspekte subsumiert werden. Auch wenn manche Forscher die Uneindeutigkeit des Konzepts kritisieren, sehen wieder andere Wissenschaftler gerade in der Abstraktheit des Begriffs einen großen Vorteil: „Als übergreifendes Konzept [trägt der Stressbegriff] zum Verständnis eines großen Bereiches von Phänomenen [bei], die die individuelle Konfrontation mit Problemen des alltäglichen Lebens und deren Bewältigung begleiten“ (Jerusalem, 1990, S. 1). Zur Ordnung der vielfältigen theoretischen Konzeptionen werden heute in Anlehnung an eine ursprünglich von Laux (1983) vorgeschlagene Taxonomie drei theoretische Stresskonzeptionen unterschieden: 1. *reaktionsbezogene*, 2. *reiz- beziehungsweise situationsbezogene* und 3. *relationale beziehungsweise interaktionistische* oder auch *transaktionale Ansätze* (z. B. Beyer & Lohaus, 2007; Jerusalem, 1990). Während Vertreter der reaktionsbezogenen Stresskonzeptionen die in der Person ablaufenden Prozesse, also die Stressreaktionen, in den Mittelpunkt des wissenschaftlichen Interesses stellen und Stress als abhängige Variable betrachten, fokussieren Vertreter der reizbezogenen Ansätze auf Stressauslöser, die Stressoren, und betrachten Stress somit als unabhängige Variable. Das transaktionale Verständnis von Stress integriert die reaktions- und reizbezogenen Ansätze und berücksichtigt darüber hinaus die Person-Umwelt-Interaktion (Beyer & Lohaus, 2007; Schwarzer, 2000). Nachfolgend werden die wichtigsten Vertreter der drei Ansätze überblicksartig vorgestellt und das transaktionale Stressmodell von Lazarus und Folkman (1984), das dem in dieser Arbeit verwendeten Stressverständnis zugrunde liegt, erläutert.

#### ***Reaktionsbezogene Stresskonzeption***

Im Fokus der reaktionsbezogenen Stresskonzeptionen stehen die physiologischen Reaktionen in Folge von Anforderungen (Thaker & Verma, 2014). Stress wird definiert als ein „psychophysischer Zustand, bei dem Abweichungen von der Homöostase vorliegen, die durch die verfügbaren, routinemäßigen Reaktionen nicht kompensiert werden können“ (Kaluza, 2011, S. 15). Dieses Verständnis geht auf den Arzt und Chemiker Selye (1936) zurück, der im Tierbereich die physiologischen Reaktionen auf bestimmte physikalische Reize wie Lärm, Hitze und Verletzungen empirisch untersucht und diese dann auch auf den Menschen übertragen hat (The American Institute of Stress, 2013). Stress ist hiernach



ein unspezifisches allgemeines Adaptionssyndrom (Taché & Selye, 1985), das unabhängig davon auftritt, ob es sich bei den Stressauslösern, in Folge derer der adaptive Prozess in Gang gesetzt wird, um positiv oder negativ besetzte Reize handelt (Beyer & Lohaus, 2007). Auf unterschiedlichste Reize reagiert der Mensch also mit einem einheitlichen, reizunspezifischen Muster an physiologischen Reaktionen: Unterstützt durch eine höhere Aktivität der Hypothalamus-Hypophysen-Nebennierenrinden-Achse kommt es zu einer verstärkten Kortisol- und Adrenalinausschüttung (Nater, Ditzen & Ehlert, 2011). Dadurch verbessern sich beispielsweise die Durchblutung sowie die Sauerstoffversorgung und die Herzfrequenz steigt an. Das Immunsystem wird kurzfristig gestärkt und die Schmerztoleranz erhöht (Beyer & Lohaus, 2007; für eine ausführliche Darstellung der körperlichen Prozesse siehe auch Kaluza, 2011). Bei dem beschriebenen Adaptionssyndrom handelt es sich vermutlich um ein altes phylogenetisches Muster an Reaktionen, das den Körper darauf vorbereitet, in bedrohlichen Situationen optimal zu reagieren und für Flucht oder Kampf bereit zu sein. Die Stressreaktionen sind also zunächst einmal Schutzfaktoren, die bei der Bewältigung spontaner Belastungen helfen. Halten die Reaktionen jedoch aufgrund von dauerhaften Belastungen an und werden die Stressreaktionen nicht mehr regelmäßig von Phasen der Entspannung unterbrochen, verliert der Körper die Fähigkeit zur Selbstregulation des Aktivierungsniveaus, sodass er schließlich auch in minder belasteten Zeiten mit Stressreaktionen reagiert. Eine solche dauerhafte Stressreaktion des Körpers kann sich dann auch negativ auf die psychische und physische Gesundheit auswirken (Beyer & Lohaus, 2007; Kaluza, 2011; siehe auch Abschnitt 2.4.3). Die Annahme über die Unspezifität der Stressreaktion ist inzwischen bereits widerlegt worden. So konnte nachgewiesen werden, dass sich Personen in ihren physiologischen Stressreaktionen durchaus unterscheiden und sich je nach mit dem Stressauslöser verbundener vorherrschender Emotion (z. B. Ärger, Angst oder Hilflosigkeit) ganz spezifische – sozusagen emotions-typische – physiologische Prozesse bestimmen lassen (Kaluza, 2011).

### ***Reizbezogene Stresskonzeption***

Vertreter der reizbezogenen Stressansätze untersuchen vor allem die Stressoren. Ein solches reizbezogenes Stressverständnis liegt beispielsweise der Lebensereignisforschung zugrunde (Beyer & Lohaus, 2007). Hiernach werden kritische Lebensereignisse wie der Tod eines Angehörigen, Krankheit, Heirat, Geburt eines Kindes oder Umzug als Stressoren betrachtet. Diesen Ereignissen ist gemeinsam, dass sie ein hohes Ausmaß an Adaption erfordern und zu psychischen und oder physischen Beeinträchtigungen führen können (Beyer

& Lohaus, 2007; Jerusalem, 1990; Kaluza, 2011). Die Beispiele Heirat und Geburt eines Kindes verdeutlichen, dass kritische Lebensereignisse nicht zwangsläufig negativ bewertet werden, sondern auch positiv bewertete Ereignisse einschneidende Veränderungen im Leben nach sich ziehen können. Holmes und Rahe (1967) haben die „Social Readjustment Scale“ entwickelt, die 43 solcher ‚live events‘ auflistet. Die kritischen Lebensereignisse wurden von den Autoren auf der Basis empirischer Befunde hinsichtlich ihres Schweregrads gewichtet, sodass sich für jede Person ein Summenwert der derzeitigen Belastungen berechnen lässt. Der Ansatz wurde vielfach kritisiert, da er die interindividuellen Unterschiede in der Anpassung auf solche kritischen Lebensereignisse unberücksichtigt lässt und nicht zwischen negativen und positiven Ereignissen unterscheidet (Jerusalem, 1990). Zudem belegen inzwischen empirische Untersuchungen zur Bedeutung von alltäglichen, kleineren Schwierigkeiten und Freuden („daily hassles and uplifts“), dass diese bessere Prädiktoren für das psychische und körperliche Wohlbefinden darstellen als die aufgrund von kritischen Lebensereignissen ermittelte Gesamtbelastung (z. B. DeLongis, Coyne, Dakof, Folkman & Lazarus, 1982; Kanner, Coyne, Schaefer & Lazarus, 1981).<sup>10</sup>

### ***Transaktionale Stresskonzeption***

Die relationalen, interaktionistischen beziehungsweise transaktionalen Stressansätze fokussieren auf die Person-Umwelt-Interaktion und berücksichtigen auch die Ressourcen einer Person. Das bekannteste Modell ist das transaktionale Stressmodell von Lazarus und Folkman (1984). Dabei handelt es sich um ein übergreifendes Modell, das reiz- und reaktionsbezogene Stressansätze vereint und diese um Bewertungsprozesse erweitert. Die Autoren definieren Stress wie folgt: „Psychological stress is a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being“ (Lazarus & Folkman, 1984, S. 19). Stress entsteht also bei einem empfundenen Ungleichgewicht zwischen Anforderungen einer Situation und den Ressourcen einer Person (Klein-Heßling & Lohaus, 2012). Somit sind nicht mehr allein äußere Reize für das Stressgeschehen verantwortlich, sondern auch kognitive Bewertungen und Bewältigungskompetenzen der Person, die wiederum auf die Situation zurückwirken können (Beyer & Lohaus, 2007). Entscheidend sind dabei drei

---

<sup>10</sup> Auch in der Arbeits- und Organisationspsychologie werden vorrangig reizbezogene Stresskonzeptionen verfolgt. Da diese in der vorliegenden Arbeit jedoch keine Rolle spielen, werden die arbeits- und organisationspsychologischen Stressansätze hier nicht näher erläutert. Ein Überblick über die Konzeptionen findet sich beispielsweise bei Ulich (2011), bei Frieling und Sonntag (1999) oder auch bei Rohmert und Rutenfranz (1975).

Formen von Bewertungsprozessen, die sogenannte *primäre Bewertung*, die *sekundäre Bewertung* und die *Neubewertung*:

Bei der primären Bewertung wird eine Situation zunächst hinsichtlich ihrer subjektiven Bedeutsamkeit für das eigene Wohlbefinden bewertet. Die Bewertung kann dabei *angenehm positiv*, *irrelevant* oder *stressrelevant* ausfallen (Beyer & Lohaus, 2007; Jerusalem, 1990; Klein-Heßling & Lohaus, 2012; Lazarus & Folkman, 1984; Lohaus et al., 2007). Anforderungen, für deren Bewältigung eindeutig ausreichend Ressourcen zur Verfügung stehen oder Ereignisse, die per se mit positiven Gefühlen verbunden sind, werden als angenehm positiv bewertet. Ein Ereignis, das keinerlei Bedeutung für das Wohlergehen einer Person hat, wird als irrelevant eingeschätzt und kann Gleichgültigkeit auslösen. Werden eine Situation und deren erfolgreiche Bewältigung hingegen als persönlich bedeutsam erlebt, handelt es sich um eine stressrelevante Einschätzung, die drei Ausprägungsformen annehmen kann: die Bewertung einer Situation als *Herausforderung* (1), als *Bedrohung* (2) oder als *Verlust* beziehungsweise *Schädigung* (3) (Lazarus & Folkman, 1984).<sup>11</sup> Die Ankündigung einer Klassenarbeit kann beispielsweise als Herausforderung betrachtet werden und würde in diesem Fall als Chance wahrgenommen, die eigenen Fähigkeiten unter Beweis stellen zu können. Sehr viel häufiger jedoch wird die Ankündigung einer Klassenarbeit als Bedrohung betrachtet und ein negativer Ausgang, beispielsweise eine schlechte Note, befürchtet (Beyer & Lohaus, 2007; Lohaus et al., 2007). Eine bereits eingetretene Beschädigung des Wohlbefindens, z. B. die Scheidung der Eltern, würde als Verlust interpretiert (Lohaus et al., 2007). In diesem Sinne beinhaltet die Bewertung eines Ereignisses als Herausforderung die Hoffnung auf einen positiven Situationsausgang und die Erwartung, hierfür die eigenen Bewältigungsfähigkeiten nutzen zu können. Auch eine Bedrohungseinschätzung richtet sich in der Regel auf zukünftige Ereignisse, entsteht jedoch in Folge der Befürchtung eines negativen Ereignisausgangs und damit verbundenen antizipierten Verlusten.<sup>12</sup> Anders als die Herausforderungs- und Bedrohungseinschätzungen beinhaltet die Bewertung einer Situation als Verlust oder Schädigung bereits eingetretene beziehungsweise nicht mehr abwendbare Beeinträchtigungen (Lazarus & Folkman, 1984). Zu beachten ist, dass Situationen mehrdeutig sein können und die stressbezogenen Bewer-

---

<sup>11</sup> Lazarus und Folkman (1987) haben die drei Ausprägungsformen der stressrelevanten Situationseinschätzung später noch ergänzt um den *Nutzen* („benefit“) und um *Attributionsprozesse* („information“), die Auskunft über die vermuteten Ereignisursachen geben. Diese werden jedoch bei der Rezeption des Modells in der heutigen Literatur nur selten berücksichtigt.

<sup>12</sup> Eine ähnliche Unterscheidung findet sich auch bei Selye (1978), der zwischen positivem und negativem Stresserleben (*Eustress* und *Distress*) unterscheidet.

tungen einer Anforderung als Herausforderung, Bedrohung oder Verlust sich abwechseln oder gar überschneiden können (Jerusalem, 1990).

Wenn die primäre Bewertung stressrelevant ausfällt, erfolgt im Rahmen der sekundären Bewertung eine Einschätzung der zur Bewältigung der Situation vorhandenen Ressourcen (Lazarus, 1999; Lazarus & Folkman, 1984):

*„[Secondary Appraisal] is a complex evaluative process that takes into account which coping options are available, the likelihood that a given coping option will accomplish what it is supposed to, and the likelihood that one can apply a particular strategy or a set of strategies effectively“ (Lazarus & Folkman, 1984, S. 35).*

Die beiden Wahrscheinlichkeiten in dieser Definition greifen die bereits beschriebene Ergebniserwartung und auch die Kompetenz- beziehungsweise Selbstwirksamkeitserwartung nach Bandura auf (1977; Bandura, 1982; Jerusalem, 1990). Stressbezogen fällt die sekundäre Bewertung aus, wenn die vorhandenen Ressourcen für die erfolgreiche Bewältigung der Situation als unzureichend eingeschätzt werden (Beyer & Lohaus, 2007) oder das Zutrauen fehlt, die zur Bewältigung geeigneten Handlungen auch ausführen zu können (Jerusalem, 1990). Ressourcen können materieller (z. B. technische Hilfsmittel, die zur Verfügung stehen), sozialer (z. B. Freunde), physischer (z. B. Gesundheit) und auch psychischer (z. B. Selbstwirksamkeitserwartung, Selbstregulationsfähigkeiten) Natur sein (Beyer & Lohaus, 2007). Die Begriffe primär und sekundär erwecken den Anschein als sei die primäre Bewertung wichtiger als die sekundäre oder als würde mit den Begriffen eine zeitliche Reihenfolge beschrieben. Eine solche Konnotation der Begriffe ist jedoch nicht intendiert. Vielmehr handelt es sich im Hinblick auf die Stressentwicklung um gleichgewichtige Bewertungsprozesse, die sich zeitlich überlappen oder auch gleichzeitig auftreten können (Jerusalem, 1990; Lazarus & Folkman, 1984). Beide Bewertungsprozesse sind „in einem transaktionalen Sinne miteinander verbunden“ (Jerusalem, 1990, S. 12).

Zu Neubewertungen der Situation kommt es schließlich immer dann, wenn sich die Ausgangssituation verändert hat oder neue Informationen vorhanden sind, die eine Neubewertung erfordern (Beyer & Lohaus, 2007). Die Neubewertungen unterscheiden sich somit qualitativ nicht von der primären und sekundären Bewertung, sie treten lediglich später auf und verdeutlichen, dass die Person-Umwelt-Interaktion ständigen Veränderungsprozessen ausgesetzt ist (Beyer & Lohaus, 2007; Jerusalem, 1990). Neben den verschiedenen geschilderten Bewertungsprozessen, beschreiben Lazarus und Folkman (1984) noch eine weitere wesentliche Komponente des Stressgeschehens, die sogenannte *Stressbewältigung*

(*coping*). Diese resultiert aus den Bewertungen und Stresseinschätzungen und dient der Reduktion des Stresserlebens (Beyer & Lohaus, 2007; Lazarus & Folkman, 1984; Schwarzer, 2000). Auf die Stressbewältigung wird unter Abschnitt 2.4.4 gesondert eingegangen.

Das transaktionale Stressmodell wurde inzwischen mehrfach erweitert und in vielen Anwendungskontexten überprüft (Beyer & Lohaus, 2007).<sup>13</sup> Es bildet die theoretische Grundlage für zahlreiche Stresspräventionstrainings im Kindes- und Jugendalter (im deutschsprachigen Raum z. B. Beyer & Lohaus, 2006; Hampel & Petermann, 2003; Klein-Heßling & Lohaus, 2012) und liegt auch dem in dieser Arbeit verwendeten Stressbegriff zugrunde.

## 2.4.2 Entstehungsbedingungen von Stresserleben

Bereits in der Einführung dieses Kapitels wurden für das Kindes- und Jugendalter zentrale Stressoren benannt. Als übergeordnete Kategorien für die verschiedenen Stressoren wurden im Abschnitt zu den reizbezogenen Stresskonzeptionen (2.4.1) bereits kritische *Lebensereignisse* und *alltägliche Schwierigkeiten und Freuden (daily hassles und uplifts)* unterschieden (Jerusalem, 1990). Neben diesen beiden sind außerdem *Entwicklungsprobleme* (Lohaus, 2009; Lohaus et al., 2007) oder auch sogenannte *normative Stressoren* (Beyer & Lohaus, 2007; Suldo, Shaunessy, Thalji, Michalowski & Shaffer, 2009) typische Stressquellen für das Kindes- und Jugendalter.

Beispiele für kritische Lebensereignisse, die im Kindes- und Jugendalter auftreten können, sind der Tod eines nahen Angehörigen, die Trennung der Eltern, ein Wohnortwechsel oder auch eine chronische Krankheit. Empirische Studien belegen jedoch, dass die alltäglichen, wiederkehrenden Stressoren (z. B. Geschwisterrivalitäten oder Mobbing in der Schule) bessere Prädiktoren für das Stresserleben darstellen als die in der Regel einmalig auftretenden kritischen Lebensereignisse, die nach einer Phase großer Verunsicherung häufig durch eine angemessene Bewältigung überwunden werden. Alltägliche Stressoren ereignen sich regelmäßig über eine längere Zeit, sodass sie häufig zu einem scheinbar nicht zu bewältigenden Stressor werden und eine dauerhafte, chronische Stressreaktion nach sich ziehen können (Beyer & Lohaus, 2007; Lohaus et al., 2007).

---

<sup>13</sup> Eine theoretische Weiterentwicklung des transaktionalen Stressmodells ist beispielsweise die sogenannte „Theorie der Ressourcenerhaltung“ von Hobfoll (1988, 1998). Die Theorie basiert auf der Annahme, dass die Wahrung der eigenen Ressourcen ein zentrales dem Menschen innewohnendes Bedürfnis ist und Stress dann entsteht, wenn zentrale eigene Ressourcen bedroht werden, bereits verloren sind oder das Bemühen um neue Ressourcen keine Erfolge erzielt. Unterschieden werden „Objektressourcen (z. B. Kleidung, eigenes Haus), Bedingungsressourcen (z. B. Gesundheit, Alter), persönliche Ressourcen (z. B. Fähigkeiten, Eigenschaften) und Energieressourcen (z. B. Geld, Zeit)“ (Beyer & Lohaus, 2007, S. 23). Eine ausführliche Darstellung der Theorie findet sich bei Hobfoll (1988, 1998).

Die für das Kindes- und Jugendalter typischen Entwicklungsaufgaben, wie z. B. der Eintritt in die Schule, der Wechsel auf eine weiterführende Schule, die Pubertät oder die Entwicklung von intimen Beziehungen, können Entwicklungsprobleme mit sich bringen und dadurch ebenfalls eine Quelle des Stresserlebens sein (Beyer & Lohaus, 2007; Lohaus et al., 2007; Suldo et al., 2009). Nach dem Übergang auf die weiterführende Schule ist z. B. ein besonders hohes Ausmaß des Stresserlebens zu verzeichnen (Hampel, Jahr & Backhaus, 2008). Auch Zukunftsängste und Sorgen über mögliche Berufs- und Karrierewege können sich im höheren Jugendalter als entwicklungsbedingte Stressauslöser erweisen (Escher & Seiffge-Krenke, 2013; Seiffge-Krenke, 2008).

Durch die alltäglichen Stressoren werden Kinder und Jugendliche besonders stark belastet (Beyer & Lohaus, 2007). Diese können sich innerhalb der Familie, in Freundschaften oder auch im Rahmen der Freizeitgestaltung abspielen. Unter allen für das Kindes- und Jugendalter stress-relevanten Bereichen kommt dem Kontext Schule eine besondere Bedeutung zu (Anda et al., 2000; Backhaus, Petermann & Hampel, 2010; Fridrici & Lohaus, 2009; Klein-Heßling & Lohaus, 2012; Yankey & Biswas, 2012). Die Reduktion schulbezogener Stressoren ist auch ein zentrales Ziel des FOSIS2-Projekts. Deswegen werden nachfolgend die schulischen Stressoren gesondert dargestellt.

### ***Alltägliche Stressoren in der Schule***

Der Kontext Schule bietet zahlreiche mögliche Stressauslöser, denen Kinder und Jugendliche täglich ausgesetzt sind: Prüfungen, überhöhte Lern- und Leistungserwartungen (Anda et al., 2000; Suldo et al., 2009), Sorgen über die Noten, Überforderung durch Hausaufgaben (Seiffge-Krenke, 2008), Zeugnisse, die Menge des Lernstoffs, Aggressionen unter Mitschülern, ein schlechtes von Konkurrenz und mangelnder Kooperation geprägtes Klassenklima (Seiffge-Krenke, 2007). Laut der zuletzt 2009/2010 durchgeführten Studie „Health Behaviour in School-aged Children (HBSC)“<sup>14</sup> fühlt sich ein Viertel der in Deutschland befragten Jugendlichen einigermaßen stark oder sogar sehr stark durch die schulischen Anforderungen belastet (HBSC Team Deutschland, 2011a). Eng gefolgt von den Hausaufgaben (Beyer & Lohaus, 2007; Klein-Heßling & Lohaus, 2012) sind die Klassenarbeiten die im Kontext Schule am häufigsten berichteten Stressoren (Beyer & Lohaus, 2007). Dies ist insofern nachvollziehbar, als dass eine bestandene oder nicht bestandene

---

<sup>14</sup> In der zuletzt durchgeführten HBSC-Studie wurde erstmalig eine bundesweit repräsentative Stichprobe aus Fünft-, Siebt- und Neuntklässlern gezogen. Diese wurden zu ihrer Gesundheit und zum Gesundheitsverhalten befragt (HBSC Team Deutschland, 2010).

Prüfung wesentliche Auswirkungen auf die künftige Entwicklung, den Schulerfolg und damit verbundene Zugangsmöglichkeiten zur akademischen und beruflichen Weiterbildung haben kann (Seiffge-Krenke, 2008). Der während einer Klassenarbeit erlebte Stress oder auch der Stress in anderen Prüfungssituationen wie das Lösen einer Aufgabe vor der Klasse an der Tafel hängt einerseits von Vorerfahrungen in ähnlichen Situationen und von den Erwartungen an die eigenen Leistungen ab. Andererseits kann auch die Art der Gestaltung einer solchen Unterrichtssituation seitens der Lehrkraft das Ausmaß des Stresserlebens beeinflussen: So verstärken z. B. das Exponieren schlechter Leistungen einzelner Schüler vor der Klasse, allgemein häufiges Strafen und Tadeln sowie selten ausgesprochenes Lob für gelungene Lernprozesse und gute Leistungen den schulbezogenen Stress (Lohaus et al., 2007). Eine vorherrschende Leistungszielorientierung seitens der Lehrperson fördert Konkurrenz unter den Schülern und trägt zu einem Wettbewerbsklima bei, das den Leistungsdruck auf den Einzelnen erhöht und somit wiederum Quelle für Stresserleben sein kann. Allgemein wirken sich ein negatives Klassenklima (Lohaus et al., 2007), eine mangelnde materielle Ausstattung des Klassenraums sowie eine schlechte Stundenplanorganisation (Suldo et al., 2009) ungünstig auf die wahrgenommene Stressbelastung aus.

### 2.4.3 Auswirkungen und Korrelate von Stresserleben

Auch wenn nicht alle Kinder und Jugendlichen angesichts von Stressoren Stresssymptome zeigen oder gar in Folge von Stresserleben gesundheitlich beeinträchtigt sind (Hampel, 2007), lassen sich doch häufige im Zusammenhang mit Stress auftretende Beeinträchtigungen ermitteln, die als Stressauswirkungen oder -folgen bezeichnet werden. Die *kurzfristigen* Stresssymptome, die zusammengenommen eine körperliche Aktivierung verursachen, wurden bereits im Abschnitt 2.4.1 erläutert. Bei dauerhaftem Stresserleben kann es zu einem wiederholten oder regelmäßigen Auftreten von Stresssymptomen kommen. Die Stressreaktionen sind dann nicht mehr situationsspezifisch, sondern situationsübergreifend und können sich chronifizieren und *langfristige* Beeinträchtigungen nach sich ziehen (Lohaus et al., 2007). Chronische Stresssymptome wirken sich nicht nur negativ auf die physische und psychische Gesundheit aus (Hampel et al., 2008), sondern bergen die Gefahr, selbst wieder zu neuen Stressoren zu werden (Lohaus et al., 2007). Die verschiedenen langfristigen Stressfolgen lassen sich drei Ebenen zuordnen: der *physiologisch-vegetativen* (körperliche Stresssymptome), der *kognitiv-emotionalen* (belastende Gedanken und Gefühle) und der *verhaltensbezogenen* (Hektik, Konzentrationsschwierigkeiten, Veränderungen im Sozialverhalten) (Beyer & Lohaus, 2007; Klein-Heßling & Lohaus, 2012; Lohaus,

2009; Lohaus et al., 2007). Die kognitiv-emotionalen und verhaltensbezogenen Stressreaktionen können zu psychischen Stressfolgen zusammengefasst werden (Lohaus et al., 2007). Nachfolgend werden deshalb physische und psychische Stressauswirkungen unterschieden.

### ***Physische Stressauswirkungen***

Dauerhaftes Stresserleben und chronische Stressreaktionen machen sich bei Kindern und Jugendlichen auf der körperlichen Ebene häufig durch Müdigkeit, Erschöpfung, Kopf- und Bauchschmerzen bemerkbar (Beyer & Lohaus, 2007). Laut aktuellen Ergebnissen der „Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland (KiGGS)“<sup>15</sup> (Robert Koch Institut (RKI, 2015) schätzen 88 % der befragten 11- bis 17-Jährigen ihren allgemeinen Gesundheitszustand als gut bis sehr gut ein. Auch in der HBSC-Studie (HBSC Team Deutschland, 2011b) zeigen sich Kinder und Jugendliche als die gesündeste Bevölkerungsgruppe. Die Ergebnisse der HBSC-Studie zu Prävalenzen von körperlichen Beschwerden im Kindes- und Jugendalter weisen jedoch auf einen Handlungsbedarf zur Wahrung und Verbesserung des kindlichen und jugendlichen Wohlbefindens hin: Aufgerundet gaben 27 % der befragten Mädchen und 15% der Jungen an, in den letzten sechs Monaten unter mindestens zwei Beschwerden pro Woche zu leiden: Am häufigsten werden Einschlafprobleme genannt (gerundet 18 % der befragten Jugendlichen), gefolgt von Kopf- (13 %), Rücken- (12 %) und Bauchschmerzen (8 %). Im höheren Jugendalter berichten Mädchen deutlich häufiger über körperliche Belastungen als Jungen (HBSC Team Deutschland, 2011b). Zu beachten ist, dass die Beschwerden nicht im Zusammenhang mit Stresserleben erfasst wurden. In anderen Studien konnte aber nachgewiesen werden, dass das Ausmaß des Stresserlebens positiv mit der Anzahl genannter Stresssymptome korreliert. Somit kann davon ausgegangen werden, dass zumindest ein Teil der berichteten körperlichen Symptome im Zusammenhang mit dem Stresserleben steht (Beyer & Lohaus, 2007; Lohaus et al., 2007).

### ***Psychische Stressauswirkungen***

Die Befunde zu langfristigen psychischen Stressfolgen sind nicht immer konsistent. Grundsätzlich lässt sich jedoch ein negativer Einfluss von alltäglichen Stressoren auf die psychische Gesundheit festhalten. Insbesondere die schulbezogenen Stressoren scheinen die psychische Entwicklung ungünstig zu beeinflussen. Anhaltender Schulstress wirkt sich

---

<sup>15</sup> KiGGS ist eine Langzeitstudie, durch die regelmäßig repräsentative Daten zur aktuellen gesundheitlichen Lage von Kindern und Jugendlichen unter 18 Jahren zur Verfügung gestellt werden (Robert Koch Institut (RKI, 2015).



negativ auf die Motivation, das Lern-, Leistungs- und Sozialverhalten aus und kann die Entwicklung der eigenen Persönlichkeit behindern (Hampel et al., 2008). Die negativen psychischen Auswirkungen sind nicht nur abhängig von den Stressoren, sondern auch von der Form der Stressbewältigung (verschiedene Formen der Stressbewältigung werden im Abschnitt 2.4.4 erläutert). Reagieren Kinder und Jugendliche auf stressreiche Situationen beispielsweise mit physischer oder verbaler Aggressivität oder mit sozialem Rückzug, kann dies die unmittelbaren Sozialbeziehungen beeinträchtigen (Lohaus et al., 2007).

Insgesamt scheinen die psychischen Stressreaktionen noch häufiger aufzutreten als die physischen (Lohaus et al., 2007). In einer Studie von Lohaus, Beyer und Klein-Heßling (2004) wurden 1699 Fünft- bis Zehntklässler zu ihrem emotionalen Befinden befragt. Am häufigsten berichteten die Kinder und Jugendlichen über Verärgerung und Wut: Von den Befragten gaben 53 % an, mehrmals pro Woche, und 5 % sogar mehrfach am Tag verärgerter zu sein. Gerundete 48 % der Kinder und Jugendlichen fühlten sich mehrmals pro Woche und 5 % mehrfach am Tag wütend (Lohaus et al., 2007). Die Daten der zuletzt durchgeführten HBSC-Studie belegen, dass von den Elf- bis Fünfzehnjährigen etwa ein Fünftel eines jeden Jahrgangs auffällige emotionale und verhaltensbezogene Probleme aufweist. Wie bei den körperlichen Beschwerden, steigen die emotionalen Probleme bei Mädchen mit dem Alter an (Bilz, Ottava & Ravens-Sieberer, 2013). Darüber hinaus konnte ein Zusammenhang zwischen Stresserleben und psychischen Beschwerden wie Angst, Depression, Aggressivität und Verhaltensproblemen nachgewiesen werden. Zudem wächst bei einem dauerhaften Stresserleben die Wahrscheinlichkeit für riskantes Gesundheitsverhalten wie z. B. Substanzmissbrauch (Jerusalem, 2006; Kraag, Zeegers, Kok, Hosman & Abu-Saad, 2006).

Die Geschlechterdifferenzen bei den physischen und psychischen Beschwerden müssen nicht zwangsläufig bedeuten, dass Mädchen tatsächlich stärker belastet sind als Jungen. Aufgrund der vorherrschenden Geschlechterstereotype ist die Bereitschaft, Belastungen und Beschwerden mitzuteilen bei Mädchen möglicherweise stärker vorhanden als bei Jungen. Schwächen und Beschwerden werden stereotyp nicht mit dem „Mann-Sein“ assoziiert, wohingegen die Reflexion eigener Schwächen durchaus zum Geschlechterstereotyp der Frau zu passen scheint (Lohaus et al., 2007). Auch wenn der Zusammenhang zwischen Stresserleben und negativen Auswirkungen auf die physische und psychische Gesundheit inzwischen klar belegt ist, so ist gleichzeitig zu beachten, dass Stress weder eine notwen-

dige noch eine hinreichende Bedingung für das Auftreten spezifischer Erkrankungen ist (Kraag et al., 2006).

#### **2.4.4 Ressourcen und Stressbewältigung**

Bisher wurden Stressauslöser und mögliche Stressreaktionen beschrieben. Nicht jedes Kind erlebt in anforderungsreichen Situationen Stress und entwickelt Stresssymptome. Ob ein Kind oder ein Jugendlicher Stress erlebt, hängt auch mit den Ressourcen und Stressbewältigungsfähigkeiten zusammen, die in einer spezifischen Situation zur Verfügung stehen (Backhaus et al., 2010; Hampel, 2007; Kraag et al., 2006).

##### ***Ressourcen***

Fällt im Sinne des transaktionalen Stressverständnisses die primäre Bewertung im Hinblick auf die Bedeutung einer Situation als stressrelevant aus und wird ein Ereignis als Herausforderung, Bedrohung oder Verlust bewertet, erfolgen im Rahmen der sekundären Bewertung eine Einschätzung der zur Bewältigung der Situation vorhandenen Ressourcen und darauf aufbauend dann Versuche der Stressbewältigung (Lazarus, 1999; Lazarus & Folkman, 1984). Stress entsteht bei einem empfundenen Ungleichgewicht zwischen Anforderungen einer Situation und den Ressourcen einer Person (Klein-Heßling & Lohaus, 2012). Ressourcen können materieller, sozialer, physischer und psychischer Natur sein. Materielle Ressourcen umfassen alle zur Verfügung stehenden Hilfsmittel (Beyer & Lohaus, 2007) wie beispielsweise finanzielle Mittel (Jerusalem, 1990) oder die räumliche Ausstattung (Lohaus et al., 2007). Die soziale Unterstützung in Familie und Freundschaften ist ein Beispiel für eine soziale Ressource, und physische Ressourcen sind beispielsweise Gesundheit und körperliche Fitness (Beyer & Lohaus, 2007). Als psychische Ressourcen für das Stresserleben werden häufig die Selbstwirksamkeitserwartung, die Fähigkeit zur Selbstregulation, internale Kontrollüberzeugungen und ein hoher Selbstwert genannt (Beyer & Lohaus, 2007; Jerusalem, 1990; Klein-Heßling & Lohaus, 2012). Der Selbstwirksamkeitserwartung kommt dabei eine besonders hohe Bedeutung zu (Hampel, Kümmel & Meier, 2009; Kraag et al., 2006; Lohaus, Fridrici & Maass, 2009; Schwarzer, 2000). Sie spielt im gesamten transaktionalen Stressprozess, also sowohl bei den verschiedenen Bewertungsprozessen als auch bei der anschließenden Stressbewältigung, eine entscheidende Rolle. So nehmen Personen mit einer hohen Selbstwirksamkeitserwartung in der Regel im Rahmen der primären Bewertung günstigere Stressbewertungen vor als Personen mit einer niedrig ausgeprägten Selbstwirksamkeitserwartung. Letztere bewerten Anforderungen angesichts

eines geringen Zutrauens in ihre eigenen Fähigkeiten häufiger als bedrohlich (Jerusalem, 1990). Auch das Ergebnis der sekundären Bewertung hängt davon ab, wie hoch das eigene Zutrauen in die Bewältigungsfähigkeiten ist (Hampel et al., 2009). Fällt das Zutrauen niedrig aus, so wird das Ungleichgewicht zwischen den Anforderungen einer Situation und den vorhandenen Ressourcen als besonders hoch wahrgenommen, wodurch sich das Stresserleben verstärkt (Jerusalem, 1990). Schließlich konnte bei hoch selbstwirksamen Personen ebenfalls eine bessere Stressbewältigung nachgewiesen werden (Hampel et al., 2009; Lyraikos, 2012; Moeini et al., 2008; Shen, 2009). Insgesamt verfolgen hoch selbstwirksame Personen ihre Bewältigungsbemühungen ausdauernder und mit mehr Anstrengung (Jerusalem, 1990).

### ***Stressbewältigung***

Zum Stressprozess zählen nach Lazarus und Folkman (1984) sowohl die im transaktionalen Modell beschriebenen Bewertungen als auch die Versuche, das Stresserleben zu verringern und zu bewältigen. Die Autoren definieren Bewältigung als kognitive und verhaltensbezogene Bemühungen einer Person, sich mit internen und externen Anforderungen auseinanderzusetzen und diese zu meistern beziehungsweise zu reduzieren. In der Literatur zur Bewältigungsforschung finden sich zahlreiche Klassifikationen zur Einordnung und Kategorisierung verschiedener Bewältigungsstrategien (z. B. Billings & Moos, 1982; Kaluza, 2011; Perez & Reicherts, 1992; Rothermund & Brandstädter, 1997). Die am häufigsten rezipierte Klassifikation von Bewältigungsverhalten ist die Unterscheidung zwischen *problem-* und *emotionsorientierter* Bewältigung (Lazarus & Folkman, 1984; Lazarus & Launier, 1978). Beide Orientierungen erfüllen unterschiedliche Funktionen: Verhaltensbemühungen zur Veränderung des Problems und die Regulation der mit dem Problem und dem Stresserleben einhergehenden Emotionen (Lazarus & Folkman, 1984). Zu den problemorientierten Bewältigungsstrategien zählen beispielsweise die Anpassung des eigenen Verhaltens an die Problemsituation oder auch die Schaffung von Kompromissen und damit verbundene Neubewertungen der Situation. Solche Strategien können unter anderem durch die Beobachtung von Modellen erlernt und auch geübt werden (Lohaus et al., 2007). Problemorientierte Bewältigung ist nicht immer funktional. Insbesondere in Situationen, die nicht mehr veränderbar scheinen oder über die keine Kontrolle ausgeübt werden kann, sind emotionsorientierte Strategien besser zur Verarbeitung geeignet (Lazarus & Folkman, 1984). Zu den emotionsorientierten Bewältigungsstrategien zählen verschiedene Strategien der Entspannung wie klassische Entspannungstechniken, aber auch das regelmäßige Einle-

gen von Ruhepausen, Spielen und Toben sowie Weinen. All diese Strategien wirken dem Stresserleben entgegen und können damit verbundene Gefühle wie Angst, Wut, Ärger, Enttäuschung und Traurigkeit verringern. Nicht immer lässt sich eine Bewältigungsstrategie eindeutig entweder der problem- oder der emotionsorientierten Bewältigung zuordnen. Dies wird deutlich am Beispiel der Suche nach sozialer Unterstützung bei Familienangehörigen, Freunden und Lehrern, die insbesondere im Kindes- und Jugendalter eine häufige und zentrale Bewältigungsstrategie darstellt: Zum einen kann sie zum Ziel haben, wirkliche Hilfe bei der eigentlichen Problemlösung zu erhalten und wäre damit eine problemorientierte Form der Bewältigung. Zum anderen kann die Suche aber auch einfach dem Wunsch nach Trost und Zuspruch entspringen und wäre somit eine emotionsorientierte Bewältigungsstrategie (Lohaus et al., 2007). Eine konstruktive Stressbewältigung stabilisiert die psychosoziale Anpassung und das Wohlbefinden, wohingegen destruktive Bewältigungsstrategien eher mit einem geringen psychischen Wohlbefinden einhergehen (Hampel et al., 2008). Wutanfälle, aggressive Verhaltensweisen (Lohaus et al., 2007), Resignation, Selbstbeschuldigung und katastrophisierende Gedanken sind beispielsweise Bewältigungsformen, die als destruktiv bezeichnet werden. Insgesamt kommen destruktive Stressverarbeitungsstrategien häufiger in der frühen und mittleren Adoleszenz zum Einsatz als in Kindheit oder spätem Jugendalter (Hampel et al., 2008).

#### **2.4.5 Angst und Stresserleben als verwandte Konstrukte**

Angst geht häufig auch mit Stresserleben einher und die beiden Konstrukte weisen somit eine gewisse Überschneidung auf. Insbesondere prüfungsängstliche Schüler erleben angesichts von Prüfungen regelmäßig Stress bei der Vorbereitung und während der Prüfung (Lohaus et al., 2007). Nach Jerusalem (1990) ist dem situativen Angst- und Stresserleben vor allem die wahrgenommene Bedrohung gemeinsam. Die kognitive Bedrohungskomponente der Zustands-Prüfungsangst ähnelt der stressbezogenen Bedrohungseinschätzung. Somit müsste die Prüfungsängstlichkeit im Sinne einer überdauernden Persönlichkeitseigenschaft vor allem geeignet sein, die Bedrohungskomponente des Stresserlebens vorherzusagen. Denn prüfungsängstliche Personen neigen dazu, Leistungsanforderungen als Bedrohung zu bewerten, so dass Leistungssituationen für diese Personen auch häufiger mit stressbezogenen Bedrohungseinschätzungen einhergehen. Die stressbezogenen Situationsbewertungen Herausforderung und Verlust hingegen finden sich in der theoretischen Angstkonzeption nicht wieder (Jerusalem, 1990).

### 2.4.6 Erfassung von Stresserleben

Wie die theoretischen Stresskonzeptionen lassen sich auch die Erhebungsinstrumente unterscheiden in Verfahren, die auf die Erfassung von Stressoren abzielen, Verfahren, die vor allem die Stressreaktionen messen, und Verfahren, die in Anlehnung an das transaktionale Stressverständnis konzipiert wurden. Neben Fragebögen zur Erfassung des Stressgeschehens existieren auch Verfahren, die auf die Messung der Stressbewältigung abzielen (Seiffge-Krenke, Gelhaar & Kollmar, 2007). Darüber hinaus gibt es Instrumente, die gleichzeitig mehrere Aspekte erheben (Eschenbeck, Lohaus & Kohlmann, 2007). Ein Beispiel hierfür ist der „Fragebogen zur Erhebung von Stress und Stressbewältigung im Kindes- und Jugendalter“ (SSKJ 3-8; Lohaus, Eschenbeck, Kohlmann & Klein-Heßling, 2006), der für Kinder und Jugendliche von der dritten bis zur achten Klasse eingesetzt werden kann. Im Fragebogen werden sowohl die individuelle Vulnerabilität für alltägliche Stressoren als auch verschiedene Bewältigungsstile und das Ausmaß der körperlichen und psychischen Stressreaktionen erfasst. Einen Überblick über die verschiedenen für das Jugendalter relevanten Instrumente geben Seiffge-Krenke, Gelhaar und Kollmar (2007). In dieser Untersuchung wird das schulbezogene Stresserleben mittels dreier Skalen erhoben, die in Anlehnung an die im transaktionalen Stressmodell von Lazarus und Folkman (1984) operationalisierten Stresseinschätzungen Herausforderung, Bedrohung und Verlust ursprünglich von Jerusalem (1990) konzipiert und im Rahmen von FOSIS2-Vorgängerprojekten an den schulischen Kontext angepasst wurden (Jerusalem, 2005a, 2008). Eine ausführliche Darstellung der Skalen findet sich unter 6.4.1.

### 2.4.7 Implikationen für die Praxis: Reduktion von Stresserleben

Bei dem in dieser Arbeit zu evaluierenden FOSIS2-Programm handelt es sich um eine schulische Intervention, durch die insbesondere die schulbezogenen Stresseinschätzungen reduziert werden sollen. Deswegen werden nachfolgend vor allem solche Ansätze zur Stressprävention und -reduktion beschrieben, die auch im Kontext Schule zum Einsatz kommen können.

#### *Maßnahmen zur Veränderung der Person und des Verhaltens*

Sowohl die primäre als auch die sekundäre Stressbewertung werden von bestimmten personalen Erwartungen und damit verbundenen Kognitionen beeinflusst: Gehen die Bewertungsprozesse in einer Situation im Sinne einer niedrigen Selbstwirksamkeitserwartung beispielsweise mit Gedanken wie „Das schaffe ich nie“, „Bei mir geht sowieso alles

schief“ einher, werden die Stresseinschätzungen ungünstiger ausfallen, als wenn mittels positiver Selbstinstruktion sogenannte „Anti-Stress-Gedanken“ wie „Ich schaffe das schon“ oder „Ich bin gut vorbereitet“ eine Situation begleiten (Lohaus et al., 2007, S. 52). Auch mithilfe der aus der kognitiven Verhaltenstherapie übernommenen Methode der kognitiven Umstrukturierung können die Stressbewertungen positiv beeinflusst werden (Hampel, 2007). Im Rahmen der kognitiven Umstrukturierung werden zunächst negative, dysfunktionale Gedanken identifiziert und danach alternative, positive, funktionale Kognitionen erarbeitet und eingeübt (Mühlig & Poldrack, 2011). Für die Ausbildung positiver Gedanken hat sich auch die sogenannte „Gedanken-Stopp-Methode“ als hilfreich erwiesen. Dabei sollen nicht enden wollende, negative Gedankenkreisläufe durch ein sich im Stillen selbst zugerufenes „Stopp“ oder „Halt“ unterbrochen und im Anschluss durch positive Anti-Stress-Gedanken ersetzt werden (Buchwald, 2011). Die Ausprägung der Stresseinschätzungen kann darüber hinaus verringert werden, indem das Kind oder der Jugendliche lernt, auch die positiven Seiten einer Situation wahrzunehmen. Der Einsatz von „Anti-Stress-Gedanken“ sowie das Erkennen positiver Seiten einer Situation können eingeübt werden (Buchwald, 2011; Lohaus et al., 2007). Dadurch werden insgesamt die Anforderungen in einer Situation eher als Herausforderung und weniger als Bedrohung bewertet (Lohaus et al., 2007). Neben solchen Kognitionsveränderungen tragen auch Lob und das Hervorheben von Stärken der Kinder und Jugendlichen sowie die Vermittlung von Erfolgserfahrungen zur Stärkung der Selbstwirksamkeitserwartung sowie des Selbstkonzepts und somit zu günstigeren Stressgedanken und positiveren Stressbewertungen bei (Lohaus et al., 2007).

Hinsichtlich der Stressbewältigung können problemlösende und emotionsregulierende Strategien unterschieden werden (Lohaus et al., 2007). Wichtigste Voraussetzung für eine kompetente Problemlösung ist die Fähigkeit, stressreiche Situationen zu identifizieren und die eigenen Stressreaktionen wahrnehmen zu können. Zudem muss die Problemsituation und das Schwierige, Stressauslösende in der Situation genau analysiert werden, um angemessene Problemlösekompetenzen einzusetzen (Lohaus et al., 2007). Beyer und Lohaus (2006) beschreiben fünf Phasen auf dem Weg zu einer gelungenen Problemlösung: 1. die Problemdefinition, in der das Problem analysiert und Ist-Soll-Vergleiche herangezogen werden; 2. die eigentliche Lösungssuche, bei der verschiedene Handlungsmöglichkeiten zur Lösung des Problems erdacht und die verschiedenen Alternativen im Sinne der sekundären Stressbewertung hinsichtlich der eigenen Fähigkeiten und Ressourcen eingeschätzt werden; 3. die Entscheidungsfindung, in der es durch das Abwägen von Vor- und Nachtei-

len der Handlungsalternativen zur Entscheidung für eine Strategie kommt, die dann 4. erprobt und 5. im Hinblick auf die ursprüngliche Problemlage bewertet wird. Diese Phasen ähneln den in Abbildung 4 (Abschnitt 4.3.6) beschriebenen Phasen des Proaktiven Handlungsregulationsmodells, das im Rahmen der FOSIS2-Maßnahmen zum Einsatz kam. Weitere zentrale problemregulierende Stressbewältigungsstrategien sind ein angemessenes Organisations- und Zeitmanagement, die Fähigkeit zur realistischen Zielsetzung sowie die Fähigkeit, eigene Grenzen zu erkennen, und wenn nötig, auf bestimmte Ansprüche zu verzichten (Buchwald, 2011).

Emotionsregulierende Strategien zur Stressreduktion setzen die Fähigkeit zur Entspannung voraus. Denn Entspannung geht mit Gefühlen wie Wohlbefinden und Gelassenheit einher, die den mit Stress verbundenen Gefühlen wie Unruhe, Angst und Anspannung konträr entgegenstehen. Die Emotionen sind dabei nicht von Kognitionen abgekoppelt. Vielmehr ist eine Person in einem emotional entspannten Zustand des Wohlfühls auch leichter in der Lage, positiv zu denken sowie die positiven Situationsanteile und die eigenen Ressourcen und Fähigkeiten wahrzunehmen als bei einem angespannten, unruhigen, ängstlichen Erleben (Buchwald, 2011). Auch die Fähigkeit zur Entspannung kann trainiert werden (Lohaus et al., 2007). Häufig eingesetzte Verfahren sind die progressive Muskelrelaxation nach Jacobson (2011), das autogene Training und Phantasiereisen. Die unterschiedlichen Ansätze der Entspannungsverfahren werden bei Lohaus, Domsch und Fridrici (2007) ausführlich beschrieben. Neben solchen direkten Entspannungstrainings können auch das einfache Hören von Musik oder das Musizieren sowie das Betreiben von Sport zur Entspannung beitragen (Lohaus et al., 2007).

### ***Maßnahmen zur Veränderung der sozialen und der Lernumwelt***

Wichtigstes Merkmal einer stressarmen Umgebung ist wohl eine gelungene Mischung aus Phasen der Belastung und Anstrengung sowie Phasen der Erholung, Ruhe und Entspannung. Ausreichender Schlaf, eine gesunde Ernährung, körperliche Bewegung und Sport, stabile Freundschaften und Beziehungen zu Gleichaltrigen und in der Familie sowie die Verfolgung eigener Interessen und das Ausbilden von Hobbys – all dies trägt zur Stressprävention und -reduktion bei. Insgesamt hat sich eine vielfältige Freizeitgestaltung, die sowohl strukturierte, angeleitete Angebote als auch Zeiten für spontane Aktivitäten, freies Spiel und Erholung erlaubt, als stresspräventiv erwiesen (Lohaus et al., 2007). Speziell in der Schule können die Stressoren schon allein durch eine angemessene räumliche Ausstattung und Gestaltung der Lernumgebung bemerkbar verringert werden (Buchwald,

2011). Ein großflächiges Schulgelände, das viele Möglichkeiten der Pausen- und Freizeitgestaltung sowie Räume für Rückzug bietet, verringert Streitigkeiten unter den Schülern, erhöht die körperliche Bewegung, verstärkt die empfundene Erholung zwischen den Unterrichtszeiten und beugt somit schulbezogenem Stresserleben vor beziehungsweise reduziert bereits aufgetretenen schulischen Stress. Die soziale Unterstützung durch Familie, Freunde und in der Schule ist eine wesentliche Ressource für das Stresserleben im Kindes- und Jugendalter. Durch konkrete Hilfestellung aber auch Trost und Zuspruch kann sie gleichermaßen problem- und emotionsregulierende Funktionen erfüllen und trägt im Sinne des Modelllernens zusätzlich eventuell noch zur Vermittlung von Bewältigungsfähigkeiten bei. Das schulbezogene Stresserleben kann auch durch ein positives Klassenklima verbessert werden. So bewirken Maßnahmen zur Förderung des Klassenklimas wie die Festlegung von Regeln und gemeinsam mit den Schülern getroffene Vereinbarungen für den Umgang mit Regelbrüchen sowie kooperative Lernformen, dass der einzelne Schüler sich als Teil eines Klassenverbandes in der Schule wohl fühlt, mehr soziale Unterstützung erfährt und dadurch auch weniger Stress erlebt. Auch ein reizarmer Arbeitsplatz in der Schule und zu Hause, in denen Ablenkungs- und Prokrastinationsmöglichkeiten durch Computer, Handy und Ähnliches vermindert werden, wirkt sich günstig auf das Stresserleben im Kindes- und Jugendalter aus (Lohaus et al., 2007).

Maßnahmen zur Veränderung der sozialen und der Lernumwelt beziehen sich auch auf das Unterrichtsverhalten der Lehrkraft. Ein Unterricht, in dem hochkonzentrierte, anforderungsreiche Phasen von solchen der Erholung mit spielerischen Elementen oder Interventionen zur Entspannung abgelöst werden, hat sich als stresspräventiv erwiesen (Lohaus et al., 2007). Insgesamt scheint sich die Anwendung vielfältiger Methoden und die Gestaltung eines abwechslungsreichen Unterrichts, in dem die Schüler individuell auch durch den Einsatz von binnendifferenzierenden Maßnahmen gefördert werden, positiv auf das schulische Stresserleben auszuwirken (Buchwald, 2011). Zudem kann auch ein Unterricht, in dem Lob und Tadel nicht nur an Leistungserfolge beziehungsweise Misserfolge geknüpft sind und wo für die Bewertung des individuellen Lern- und Leistungsfortschritts nicht nur die soziale, sondern auch die individuelle Bezugsnorm herangezogen wird, die Angst vor schlechten Noten und das damit verbundene Stresserleben verringern (Lohaus et al., 2007). Inzwischen existieren zahlreiche Stresspräventionstrainings beziehungsweise -programme für das Kindes- und Jugendalter, die mehrere der beschriebenen Einzelmaßnahmen kombinieren und häufig auf dem Stressimpfungstraining von Meichenbaum (2012) basieren



(Hampel, 2007; Hampel et al., 2009). Das Training besteht aus drei Phasen, der *Informationsphase* (1), der *Lern- und Übungsphase* (2) und der *Anwendungs- und Posttrainingsphase* (3). Die Informationsphase dient der Vermittlung von Wissen über das Stressgeschehen und der Diagnostik der Stresssymptomatik sowie der Ausbildung einer guten, stabilen therapeutischen Beziehung. In der Lern- und Übungsphase werden Fähigkeiten zur kognitiven Umstrukturierung, zur positiven Selbstinstruktion, zur Problemlösung und zur Entspannung erlernt und bereits vorhandene Bewältigungsstrategien gefördert. Die erworbenen Fähigkeiten werden dann in der Anwendungsphase auf Alltagssituationen übertragen, mittels Rollenspielen, Verhaltens- und Imaginationübungen sowie Modelllernen eingeübt und über eine stufenweise Konfrontation mit Stressoren erprobt. In diese letzte Phase fällt auch ein Rückfalltraining, im Rahmen dessen das Vorgehen bei Rückfällen in alte Reaktions- und Bewältigungsmuster besprochen wird (Meichenbaum, 2012). Für das Kindes- und Jugendalter sind im deutschsprachigen Raum vor allem drei Programme zum Stressmanagement bekannt geworden, die an das Stressimpfungstraining von Meichenbaum (2012) angelehnt sind: das „Anti-Stress-Training (AST)“ für Kinder und Jugendliche im Alter von acht bis dreizehn Jahren von Hampel und Petermann (2003), das Stresspräventionstraining „Bleib locker“ für Dritt- und Viertklässler von Klein-Heßling und Lohaus (2012) und das Stressbewältigungsprogramm „Stress nicht als Katastrophe erleben (SNAKE)“ für Acht- und Neuntklässler von Beyer und Lohaus (2006).<sup>16</sup>

#### **2.4.8 Zusammenfassung: Stresserleben**

Neben der schulischen Selbstwirksamkeit und der Prüfungsängstlichkeit wird auch das schulbezogene Stresserleben als Kriterium zur Evaluation des FOSIS2-Programms herangezogen. Die Schule ist eine besonders häufige Quelle für Stresserleben, da sich hier zahlreiche Stressoren identifizieren lassen, z. B. Prüfungen, überhöhte Lern- und Leistungserwartungen (Anda et al., 2000; Suldo et al., 2009), Sorgen über die Noten, Überforderung durch Hausaufgaben (Seiffge-Krenke, 2008), Konflikte unter Mitschülern und ein schlechtes von Konkurrenz und mangelnder Kooperation geprägtes Klassenklima (Seiffge-Krenke, 2007). Schulbezogener Stress kann sich negativ auf die Motivation, das Lern-, Leistungs- und Sozialverhalten auswirken (Hampel et al., 2008). Nach dem transaktionalen Stressverständnis von Lazarus und Folkman (1984) hängt das Stresserleben von verschiedenen Bewertungsprozessen (z. B. der primären und der sekundären Bewertung) und der

---

<sup>16</sup> Einen Überblick über die Ziele, Inhalte, Anwendungsbereiche und Effekte dieser Trainings findet sich bei Lohaus (2009).

Stressbewältigung ab. Bei der primären Bewertung wird die subjektive Bedeutung einer Situation eingeschätzt. Dieser Bewertungsprozess fällt stressrelevant aus, wenn die Situation als Herausforderung, als Bedrohung oder als Verlust eingeschätzt wird. Genau diese drei Stresseinschätzungen stehen im Fokus der vorliegenden Untersuchung und sollen mithilfe der FOSIS2-Maßnahmen günstig beeinflusst werden, wobei speziell die Wahrnehmung schulischer Anforderungen als Herausforderung gestärkt und das Erleben von schulischen Situationen als Bedrohung oder Verlust verringert werden sollen. So werden im FOSIS2-Programm beispielsweise klimaförderliche und kooperative Maßnahmen eingesetzt, um unter anderem auch das Stresserleben zu verbessern. Denn ein Schüler, der in den Klassenverband integriert ist und sich in der Schule wohl fühlt, wird auch weniger Stress erleben (Lohaus et al., 2007).

Nachfolgend werden Förderstrategien und -programme beschrieben werden, die das Ziel verfolgen, gleichzeitig mehrere motivational relevante Variablen zu verbessern.

## **2.5 Strategien und Ansätze zur Motivationsförderung in der Schule**

Lehrer aller Schulformen sehen sich immer wieder mit dem Problem konfrontiert, dass manche Schüler einer Klasse hochmotiviert und andere im Gegensatz dazu unmotiviert und apathisch erscheinen (Linnenbrink & Pintrich, 2003). Insbesondere nach dem Übergang von der Grund- zur weiterführenden Schule ist über alle Schultypen hinweg eine deutliche Abnahme von motivationalen Ressourcen festzustellen und mit Motivationsproblemen zu rechnen (Wigfield et al., 2011). Die Frage, wie die Motivation wirksam gefördert werden kann, beschäftigt deshalb seit Jahrzehnten Lehrkräfte und Motivationsforscher (Fries & Souvignier, 2009; Rheinberg & Krug, 2005b). Auch wenn die Mehrzahl der Studien der Motivationspsychologie korrelativ und laborexperimentell gestaltet ist, existieren bereits einige theoretisch fundierte, in der schulischen Praxis erprobte und evaluierte Motivationsförderprogramme, die Rückschlüsse auf praxiswirksame Förderstrategien im Anwendungsfeld Schule erlauben (Wigfield et al., 2012). Diese Strategien zur Förderung der Motivation zielen auf die Beeinflussung der individuellen motivationsbezogenen Kognitionen (z. B. Erwartungen, Wertvorstellungen, Attributionen) oder auch auf die Veränderung von lernbezogenen Emotionen (z. B. Prüfungsangst und Lernfreude) ab. Darüber hinaus können Fördermaßnahmen situative Merkmale betreffen, z. B. die Art der Lernaufgaben, Unterrichtsmerkmale und Klassen- beziehungsweise Schulstrukturen. Auch die Förderung von sozialen Beziehungen, z. B. der Schüler-Schüler- beziehungsweise Schüler-Lehrer-Beziehung, kann die Lern- und Leistungsmotivation stärken (Rheinberg, 2005; Wigfield et

al., 2011). Meist verfolgen Ansätze zur Förderung motivationaler Prozesse gleichzeitig mehrere Ziele: Sie versuchen situative Veränderungen zu erzielen und dadurch auch lern- und leistungsbezogene Kognitionen und Emotionen zu beeinflussen beziehungsweise soziale Beziehungen zu stärken. Spezifische Ansätze und Prinzipien zur Förderung der zentralen motivational relevanten Zielkriterien des FOSIS2-Programms (schulische Selbstwirksamkeitserwartung, Prüfungsängstlichkeit und schulbezogenes Stresserleben) wurden bereits beschrieben. Nachfolgend soll ein weiteres Förderprinzip etwas ausführlicher dargestellt werden: die wahrgenommene Binnendifferenzierung im Unterricht. Diese externe motivational relevante Ressource der Lernumwelt beschreibt ein Unterrichtsmerkmal, das neben den erläuterten personalen Schülerkriterien ebenfalls durch den Einsatz der FOSIS2-Strategien gefördert werden soll. Im Anschluss werden exemplarisch Interventionsstudien referiert, die verschiedene Maßnahmen zur Förderung mehrerer motivational relevanter Variablen kombinieren, darunter auch zwei sehr bekannte und vielfach zitierte Motivationsförderprogramme: das TARGET-Programm von Ames (1992b) und das Origins-Programm von DeCharms (2011).

### 2.5.1 Binnendifferenzierung

„*Binnendifferenzierung*, auch *innere Differenzierung* genannt, bedeutet die individuelle Förderung einzelner Lernender innerhalb der bestehenden heterogenen Lerngruppe durch besondere Maßnahmen“ (Friedl et al., 2010, S. 51). Statt eines gleichartigen Lernangebots für alle Schüler werden innerhalb einer Klasse oder Lerngruppe verschiedene Lernangebote eingesetzt und somit die heterogenen Voraussetzungen der Lernenden berücksichtigt (Bohl, Batzel & Richey, 2012; Bohl, Bönsch, Trautmann & Wischer, 2012), z. B. die unterschiedlichen Ausprägungen von Fähigkeiten, Vorwissen, Motivation, Zielsetzung, Interessen, Kenntnissen über Lernstrategien und sozialen Kompetenzen (Friedl et al., 2010). Ein differenzierter Unterricht folgt der Annahme, dass entsprechend der Heterogenität einer Lerngruppe nicht alle Schüler von den gleichen Lernmaterialien profitieren und „der Vielfalt (der Schüler) mit der Vielfalt (der Lehre, des Unterrichts) begegnet werden“ soll (Bohl & Bönsch et al., 2012, S. 6). Im deutschsprachigen Raum wird der inneren Differenzierung die *äußere Differenzierung* gegenübergestellt. Diese bezieht sich auf das in Deutschland (noch immer) meist gegliederte Schulsystem in Haupt- und Realschulen sowie Gymnasien. In der vorliegenden Arbeit wird die Binnendifferenzierung untersucht. Diese kann auch als eine übergeordnete Form der *Individualisierung* verstanden werden. Bei der Individualisierung wird auf der Basis zuvor diagnostizierter individueller Voraus-

setzungen der Schüler einer Lerneinheit für jeden Lernenden einzeln ein geeignetes Set an Lernmaterialien zusammengestellt. Der Begriff der Differenzierung bezieht sich auch auf gruppenbezogene Anpassungen des Lernangebots, sodass beispielsweise nicht jeder Schüler einer Klasse ein individuell seinen Lernvoraussetzungen entsprechendes Lernangebot erhält, sondern innerhalb der Klasse kleinere, hinsichtlich bestimmter Merkmale homogene Gruppen von Schülern an den gleichen Aufgaben arbeiten. Differenzierende Maßnahmen erfordern somit häufig einen geringeren Aufwand als ein passgenauer individualisierter Unterricht, der im Schulalltag auch nur schwer zu erreichen ist (Bohl & Batzel et al., 2012).

Die Umsetzung von Binnendifferenzierung im Unterricht kann auf zweierlei Weise erfolgen: Entweder weisen die Lehrer den einzelnen Schülern oder Gruppen von Schülern passende Aufgaben zu oder die Schüler selbst wählen aus einem breiten Lernangebot für sich angemessene Aufgaben aus (Bohl & Batzel et al., 2012; Kress & Pappas, 2013). Die erste Variante, die von Kress und Pappas (2013) auch *gelenkte Differenzierung* genannt wird, erfordert ein hohes Engagement der Lehrkraft, die ihre Schüler hierfür sehr gut kennen muss (Bohl & Batzel et al., 2012). In der zweiten Variante, der *natürlichen Differenzierung* (Kress & Pappas, 2013), liegt die Herstellung der Passung von Aufgaben und Fähigkeiten auf Seiten der Schüler. Diese sollten jedoch bei der Wahl der Aufgaben weiterhin durch die Lehrkraft unterstützt werden, um eine Unter- oder Überforderung zu vermeiden (Bohl & Batzel et al., 2012; Kress & Pappas, 2013). Im Unterricht können die Menge an gestellten Lernaufgaben, das erwartete Tempo bei der Aufgabenbearbeitung, die Aufgabenschwierigkeit, die Anzahl der erlaubten Lösungsversuche, das Ausmaß an Anleitung beziehungsweise Selbstbestimmung beim Lernen und viele andere Dimensionen differenziert werden (Bohl & Batzel et al., 2012). Differenzierungsmaßnahmen werden typischerweise im Rahmen von Lern- und Übungszeiten mithilfe von eher offenen Aufgabenformaten, Wochenplänen, Freiarbeit, Projektarbeit oder auch verschiedenen Formen kooperativen Lernens, wie z. B. Lernzirkel, Stationsarbeit oder Gruppenpuzzle, realisiert (Bohl & Batzel et al., 2012; Kress & Pappas, 2013). Einen Überblick sowie konkrete Umsetzungshilfen für zahlreiche Methoden zur Differenzierung geben Kress und Pappas (2013).

Die Forschungsbefunde zu den Auswirkungen eines binnendifferenzierten Unterrichts auf die Schülerleistungen sind uneinheitlich, was nach Bohl und Batzel et al. (2012) auf die unterschiedlichen Untersuchungsmethoden zurückzuführen ist. Studien zur Wirksamkeit von Binnendifferenzierung variieren hinsichtlich der zugrunde liegenden Operationalisie-

rung des Konstrukts, hinsichtlich der eingesetzten Messinstrumente, Methoden und der Zielgruppe. Zudem fehlen kontrollierte Längsschnittstudien, die Hinweise auf die Wirkmechanismen von differenzierenden Unterrichtsmaßnahmen erlauben. Auch wenn sich die Binnendifferenzierung aufgrund der uneinheitlichen Forschungslage gegenüber dem traditionellen lehrerzentrierten Unterricht nicht eindeutig als überlegen erweist (Bohl & Batzel et al., 2012), so gilt sie doch als wichtiges Unterrichtsprinzip zur Förderung der schulischen Motivation (Wigfield et al., 2011) und empirische Studien belegen einen moderaten Effekt auf die schulischen Leistungen (Schneider & Ludwig, 2012). Nach Lou, Abrami und Spence (2000) nehmen die positiven Leistungsauswirkungen zu, wenn die Differenzierung nicht nur auf der Basis der zuvor diagnostizierten Fähigkeitsniveaus, sondern auch unter Berücksichtigung anderer Dimensionen erfolgt und wenn die Lehrer in binnendifferenzierenden Maßnahmen fortgebildet werden sowie kooperative Lernformen zur Gruppenbildung einsetzen. Durch die differenzierte Zuteilung von Aufgaben im Unterricht steigt auch die Chance auf eigene Lernerfolge und somit ebenfalls die Möglichkeit, höhere Kompetenzüberzeugungen ausbilden zu können. Zu beachten ist bei dem Einsatz binnendifferenzierender Maßnahmen allerdings die Gefahr der Stigmatisierung<sup>17</sup> (Schneider & Ludwig, 2012) von Schülern, die von der Lehrkraft leichte Aufgaben erhalten. Eine offene, für alle Schüler sichtbare Zuteilung der Aufgaben, durch die einzelne Schüler exponiert werden, sollte daher vermieden werden (Kress & Pappas, 2013). Insgesamt kann eine erfolgreiche Binnendifferenzierung nur durch eine gute und intensive Vorbereitung der Lehrkraft erreicht werden, bei der auch die sozialen Dynamiken einer Klasse berücksichtigt werden (Schneider & Ludwig, 2012).

### 2.5.2 Interventionsstudien zur Motivationsförderung in der Schule

Interventionsstudien zur Motivationsförderung für Schüler und Studierende gibt es seit mehr als zwanzig Jahren (Rheinberg & Krug, 2005b).<sup>18</sup> Nachfolgend werden einige der Interventionsstudien aus dieser Zeit exemplarisch vorgestellt. Die Darstellung erhebt nicht

---

<sup>17</sup> Eine vielfach diskutierte Auswirkung der in Deutschland durch das mehrgliedrige Schulsystem vorgegebenen äußeren Differenzierung ist die soziale Stigmatisierung. Gegner werfen der Dreiteilung von Haupt- und Realschulen sowie Gymnasien soziale Ungerechtigkeit und mangelnde Chancengleichheit vor (Schneider & Ludwig, 2012). Die im Jahr 2011 begonnene Berliner Schulreform, bei der Haupt-, Real- und Gesamtschulen zu einer Schulform, der integrierten Sekundarschule, zusammengelegt werden (Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin, 2014), stellt einen Versuch dar, diese negativen Auswirkungen der äußeren Differenzierung zu verringern (Senatsverwaltung für Integration, Arbeit und Soziales Berlin, 2011).

<sup>18</sup> Eine Übersicht über Motivationsförderprogramme geben Rheinberg und Krug (2005a). Im Jahr 2007 ist in der Zeitschrift *Educational Psychologist* zudem ein Sonderheft zum Thema *Introduction to Motivation at School: Interventions that Work* (Wigfield & Wenzel, 2007) erschienen.

den Anspruch auf Vollständigkeit. Vielmehr soll die Auswahl der Studien die Breite und Vielfalt der theoretischen und praktischen Ansätze veranschaulichen.

Streblow, Schiefele und Riedel (2012) entwickelten ein „Programm zur Förderung der Lesekompetenz und der Lesemotivation (LeKOLeMO)“ für Siebtklässler. Die Maßnahmen basieren auf der Theorie der Selbstbestimmung von Deci und Ryan (1993) und zielen darauf ab, die intrinsische Lesemotivation und das Selbstkonzept der eigenen Lesefähigkeit zu steigern. So sollten die Schüler beispielsweise Lesethemen selbstbestimmt wählen können und wurden als Experten in einem Themenbereich ausgebildet, sodass sie eine Kleingruppe in einem Thema anleiten konnten. Sie erhielten zudem positive, motivationsförderliche Rückmeldungen. Das LeKOLeMO-Training wurde anhand einer Stichprobe von 235 Schülern (acht Trainings- und fünf Kontrollklassen) evaluiert. Die Ergebnisse belegen die Wirksamkeit des Trainings im Hinblick auf die Lesekompetenz. Bezüglich der motivationalen Kriterien konnte bei den Interventionsschülern verglichen mit den Kontrollschülern eine Abnahme der intrinsischen Lesemotivation verhindert und das Selbstkonzept der eigenen Lesefähigkeit erhöht werden. Die Autoren empfehlen, die Effekte künftig im Hinblick auf ihre Langfristigkeit zu überprüfen (Streblow et al., 2012).

Martin (2008) hat eine multidimensionale Intervention zur Verbesserung verschiedener motivational relevanter Faktoren (z. B. Selbstwirksamkeitserwartung, Lernzielorientierung, Wertvorstellungen, Ausdauer, Planung, Aufgabenbewältigung, Ängstlichkeit und Fehlervermeidung) entwickelt. In der Intervention wurden die Zielkriterien im Rahmen einzelner Module mit den Schülern systematisch und direkt bearbeitet: Zunächst wurde ein Grundverständnis über das jeweilige Zielkonstrukt vermittelt und im Anschluss daran sollten die Schüler die einzelnen Facetten des Konstrukts bei sich selbst beobachten. Bezogen auf die Selbstwirksamkeitserwartung wurden sie beispielsweise aufgefordert, negative Kompetenzüberzeugungen zu reflektieren und ihre eigenen Erfolge zu beobachten. In einem nächsten Schritt sollten die Schüler dann jeden negativen Gedanken bezüglich eigener Fähigkeiten und Kompetenzen durch einen positiven Gedanken ersetzen und Strategien entwickeln, wie sie selbst mehr Erfolge erfahren können. Daran anknüpfend wurden die Schüler aufgefordert festzuhalten, welche der Erfolgs-Strategien für sie persönlich am besten funktionieren. Zum Schluss sollten die Schüler die eigenen Erarbeitungen noch einmal zusammenfassend nachzeichnen, den Abschluss des Moduls kommentieren und vom Lehrer unterzeichnen lassen. Die Intervention wurde an 26 australischen High-School-Jungen erprobt und mit einer direkten Kontrollgruppe derselben Schule ( $n = 27$ ) verglichen. Das

Durchschnittsalter lag bei 15 Jahren. Zudem wurden die Werte der Interventionsschüler mit den Werten einer größeren, bereits früher erfassten Stichprobe von 3381 Schülern verglichen, die ebenfalls nur Jungen mit einem durchschnittlichen Alter von 15 Jahren umfasste. Verglichen mit der direkten Kontrollgruppe und der größeren Vergleichsgruppe zeigten sich ein bedeutsamer Anstieg in den Wertvorstellungen, der Lernzielorientierung, Planung, Aufgabenbewältigung und Ausdauer sowie eine substantielle Reduktion der Fehlervermeidung. Keine signifikanten Veränderungen fanden sich in der Selbstwirksamkeitserwartung und der Ängstlichkeit.

In einem FOSIS2-Vorgänger-Projekt evaluierten Röder und Jerusalem (2007) in 12 Klassen ( $n = 180$  Schüler) der siebten und achten Jahrgangsstufe die Effekte des in Abschnitt 2.3.6 beschriebenen Transparenzpapiers in Abhängigkeit von der Umsetzungshäufigkeit der Maßnahme. Die Ergebnisse der Studie belegen, dass sich ein häufiger Einsatz der Transparenzpapiere im Unterricht günstig auf die schulische Kompetenzüberzeugung auswirkt. Im Hinblick auf die Prüfungsängstlichkeit konnten wider Erwarten keine Effekte gefunden werden.

In Anlehnung an Epstein (z. B. 1988) beinhaltet das bekannte, von Ames (1992a) konzipierte Programm TARGET verschiedene Unterrichtspraktiken zur Förderung einer Lernzielorientierung. Das Akronym TARGET steht für sechs Dimensionen, denen diese Praktiken zugeordnet werden können. Die erste Dimension bezieht sich auf die Aufgabengestaltung („task design“): Unterrichtsaufgaben sollten das Interesse der Kinder wecken, indem sie eine persönliche beziehungsweise praktische Relevanz der Lerninhalte aufzeigen. Zudem sollten die Schüler darin unterstützt werden, sich selbst realistische Nahziele zu setzen und Lern- sowie Selbstregulationsstrategien zu erwerben, um so verstärkt eigene Lernerfolge zu erfahren. Die zweite Dimension betrifft die Verantwortungsübernahme („distribution of authority“): Schülern sollten Wahl- und Mitbestimmungsmöglichkeiten eingeräumt und Verantwortung für ihre eigenen Lernaktivitäten übertragen werden. Dadurch erleben sie sich selbst als Urheber und Gestalter der eigenen Lernhandlungen und können ihre eigenen Lernfortschritte und -erfolge bewusster erkennen. Im Rahmen der dritten Dimension, Lob und Anerkennung im Unterricht („recognition of students“), weist Ames auf die Bedeutsamkeit der Anerkennung von Anstrengungsbemühungen der Schüler hin. Jeder Schüler sollte Gelegenheiten für Lob und Belohnungen erhalten, um wiederum mehr Erfolge zu erfahren. Die vierte Dimension hebt die Bedeutung von Gruppenarbeitsphasen („grouping arrangements“) und den Einsatz von kooperativen Lernformen, bestehend aus

heterogenen Gruppen, hervor. Die Arbeit in Gruppen steigert das Zugehörigkeitsgefühl und verbessert das Sozialklima, wodurch auch die Lernprozesse positiv beeinflusst werden können. Zur Förderung der Lernzielorientierung sollte im Rahmen von Bewertungsprozessen („evaluation practices“), der fünften Dimension, eine Fehlerkultur geschaffen werden, und bei der Bewertung sollten vor allem die individuelle Verbesserung und der Fortschritt (individuelle Bezugsnorm) der Schüler berücksichtigt werden. In diesem Sinne sollten Schülern vor allem auch Lerngelegenheiten ermöglicht werden, bei denen sie sich verbessern können. Schließlich weist die sechste und letzte Dimension auf die Wichtigkeit der zur Verfügung stehenden Zeit hin („time allocation“). Schüler mit Lernschwierigkeiten sollten längere Aufgabenbearbeitungszeiten erhalten und insgesamt sollte der Unterricht den Schülern ermöglichen, die für sich selbst optimalen Bearbeitungszeiten zu planen. Auch dadurch sind eher Lernerfolge möglich. Im Rahmen des Projekts TARGET wurden Lehrkräfte angehalten, die genannten Strategien täglich in ihren Unterricht zu implementieren. Ergebnisse von Untersuchungen bis 1992 belegen, dass die Lernzielorientierung durch den Einsatz der Strategien gefördert werden kann (Ames, 1992a).<sup>19</sup>

Eines der ältesten Motivationsförderprogramme ist das Origin-Training von DeCharms (1979). Nach DeCharms (1979) entsteht eine aktuelle Motivation, wenn sich eine Person als Urheber der eigenen Handlung erlebt. In diesem Sinne führt DeCharms den Begriff der *kausalen Autonomie* als Voraussetzung für die Motivationsentstehung ein: „Der Mensch hat das Bedürfnis, Urheber von Kausalitätsabläufen, der primäre kausale Ausgangspunkt, d. h. Meister seines Verhaltens zu sein; er strebt nach Autonomie“ (DeCharms, 2011, S. 15). Das Origin-Training von DeCharms (1979) hat zum Ziel, die kausale Autonomie von Lehrern und Schülern zu fördern. Dazu wurden Lehrkräfte von fünften bis achten Jahrgängen in verschiedenen Übungen und Selbsterfahrungselementen dazu angeregt, sich in Situationen zu begeben, in denen sie sich selbst als Meister und Urheber ihrer Handlungen wahrnehmen konnten, und in Situationen, wo sie klar von anderen abhängig waren und Vorgaben befolgen mussten (z. B. Blindenführerspiel<sup>20</sup>). Die Teilnehmer wurden aufgefordert, ihre eigenen Emotionen und ihre Motivation in den unterschiedlichen Übungen zu beobachten und zu reflektieren. Durch die Intervention sollten die Lehrkräfte sich zunehmend mehr als Urheber ihrer eigenen Handlungen erleben, um dann auch im Unterricht als

---

<sup>19</sup> Eine ausführliche Beschreibung des Studiendesigns und der Ergebnisse liegt bislang nicht vor.

<sup>20</sup> Beim Blindenführerspiel werden einem Interventionsteilnehmer die Augen verbunden und ein anderer muss diesen betreuen. Dadurch ist die Person mit verbundenen Augen vollkommen abhängig von ihrer Betreuungsperson und muss ihre Anweisungen befolgen.



autonome Verhaltensmodelle für die Schüler aufzutreten. Auch die Schüler wurden in Unterrichtseinheiten in ihrer Wahrnehmung gestärkt, Urheber ihres eigenen Handelns zu sein. Ergebnisse der Programmevaluation bestätigen, dass die Schüler den Unterricht durch das Origin-Training autonomieförderlicher erleben. Da das Training aus verschiedenen, sehr komplexen Programmmodulen besteht, wurde es nie vollständig repliziert. Einzelne Programmelemente fanden aber Eingang in andere, neuere Motivationsförderprogramme (Rheinberg & Krug, 2005a) und das theoretische Konzept der kausalen Autonomie wird beispielsweise auch in der Theorie der Selbstbestimmung (Deci & Ryan, 1993) aufgegriffen.

### **2.5.3 Zusammenfassung: Strategien und Ansätze zur Motivationsförderung in der Schule**

Die beschriebenen Interventionsstudien verdeutlichen, dass Maßnahmen zur Motivationsförderung in Bildungsinstitutionen verschiedenste motivationale Ziele verfolgen können. Zudem zeigen sie, dass viele in der Praxis erprobte und evaluierte Programme nicht immer die intendierten positiven Programmwirkungen nach sich ziehen und erwartete Effekte auch ausbleiben können. Dies belegt die bereits in der Einleitung beschriebene Schwierigkeit, eindeutige, allgemein gültige Strategien zur Motivationsförderung zu entwickeln und kann einerseits als Warnung vor zu euphorischen Erwartungen bezüglich möglicher Effekte einer in der schulischen Praxis erprobten und evaluierten Intervention gesehen werden. Andererseits sind die positiven Ergebnisse der hier dargestellten Studien auch eine Ermutigung zur Fortführung von weiteren quasi-experimentellen Interventionsstudien in der schulischen Praxis (Wigfield et al., 2012), die ein interdisziplinäres und multidimensionales Vorgehen wählen und unterschiedliche theoretische Perspektiven sowie verschiedene motivational relevante Ressourcen und Risikofaktoren berücksichtigen (Martin, 2008; Pintrich, 2003). Das in dieser Untersuchung evaluierte Projekt FOSIS2 kann als eine solche Fortführung der Interventionsforschung zur Motivationsförderung betrachtet werden. FOSIS2 ist ein Versuch, theoretisch fundierte Prinzipien zur Beeinflussung von Ressourcen und Risikofaktoren für die Motivation in der schulischen Praxis zu erproben und zu evaluieren. Bevor dieses quasi-experimentelle Projekt dargestellt wird, sollen zunächst noch wesentliche Erkenntnisse der Evaluations- und Implementationsforschung überblicksartig erläutert werden. Denn Ziel der vorliegenden Untersuchung ist die Evaluation der Wirksamkeit der FOSIS2-Maßnahmen, speziell auch in Abhängigkeit von der Implementation.

### 3 Evaluations- und Implementationsforschung

Während Schnack im Jahr 1997 in einer Übersichtsarbeit zu Evaluationsstudien im deutschsprachigen Raum darauf hinweist, dass die Anzahl der Veröffentlichungen zu diesem Thema begrenzt ist, existieren inzwischen zahlreiche publizierte Evaluationsstudien. Insbesondere im Bildungswesen, speziell im Kontext von Schule und Universität, ist der Evaluationsbegriff in aller Munde (Bauer, 2007). Gollwitzer und Jäger (2009) konstatieren einen „inflationären“ Gebrauch des Evaluationsbegriffs im Bildungsbereich, wo „so ziemlich alles evaluiert wird“ (Vorwort, S. XI). Im Hinblick auf die schulische Ausbildung werden der Unterricht, ganze Schulen oder Bildungssysteme evaluiert (Gollwitzer & Jäger, 2009). Sogar die schlichte Leistungsbeurteilung im Unterricht wird als Evaluation bezeichnet und Lehrkräfte, die ihre Schüler am Ende einer Stunde nach der Zufriedenheit mit dem Unterricht fragen, nennen dies ebenfalls eine Evaluation. Viele dieser sogenannten Evaluationen charakterisieren schlichte Bewertungsprozesse, die losgelöst sind von theoretischen Annahmen über Evaluationskriterien und Wirkungszusammenhänge und somit nicht dem wissenschaftlichen Verständnis von Evaluation entsprechen (Bauer, 2007). Durch Studien wie TIMMS („Trends in International Mathematics and Science Study“; siehe z. B. Bos & Wendt, 2013; Bos, Wendt, Köller & Selter, 2012; International Association for the Evaluation of Educational Achievement, 2011; Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2014) und PISA („Programme for International Student Assessment“; OECD, 2011, 2014a, 2014b) ist aber auch das Interesse an wissenschaftlicher Evaluation, Qualitätssicherung, Weiterentwicklung und Verbesserung von Schule und Unterricht stark gewachsen (Blömeke, 2004; Helmke, 2014), und sämtliche im Bildungsbereich angesiedelten Ministerien und Institute stellen inzwischen standardisierte Manuale oder Anweisungen zur Durchführung von Evaluationen zur Verfügung (Helmke, 2014). Die Pädagogische Psychologie greift dieses Interesse auf und versucht unter Einhaltung bestimmter wissenschaftlicher Standards Maßnahmen zu konzipieren und zu evaluieren, die soziale, motivationale und emotionale Variablen verändern wollen (Köller, 2009). Ziel der vorliegenden Arbeit ist die Evaluation des FOSIS2-Programms, das soziale und motivationale Ressourcen und Risikofaktoren der Interventionsschüler günstig beeinflussen soll. Hinsichtlich des Vorgehens bei der Evaluation wird hier das traditionelle Evaluationsverständnis der Bildungsforschung aufgegriffen, wonach unter Evaluation vor allem die Untersuchung der *Wirksamkeit* eines Programms oder einer

Intervention verstanden wird (Helmke, 2014). Neben dieser klassischen Wirksamkeitsevaluation der FOSIS2-Programmeffekte soll außerdem der Einfluss des Umsetzungsverhaltens der Lehrkräfte untersucht werden. Denn die Wirksamkeit eines Programms hängt eng mit ihrer Implementation zusammen (Durlak, 2013). Die Untersuchung der Umsetzung von theoretisch fundierten Forschungsbefunden in die Praxis ist Aufgabe der *Implementationsforschung* (Petermann, 2014a). Nach einer zusammenfassenden Darstellung wesentlicher Erkenntnisse der Evaluationsforschung werden zentrale Merkmale der Implementationsforschung dargestellt.

### 3.1 Evaluationsforschung

In der Literatur zur Evaluationsforschung existiert weder ein einheitliches Evaluationsverständnis noch werden die Begriffe zur Klassifikation von Evaluationszielen, -arten, -rahmenbedingungen und anderen Dimensionen von verschiedenen Autoren konsistent verwendet (z. B. DeGEval – Deutsche Gesellschaft für Evaluation e. V., 2008; Wottawa & Thierau, 2003). Die Darstellung aller Evaluationsansätze und in der Literatur zur Evaluationsforschung referierten Begriffe würde den Rahmen dieser Arbeit sprengen. Nach einer einleitenden Begriffsdefinition werden hier überblicksartig die Funktionen, Ziele, Rahmenbedingungen, Merkmale und Standards von Evaluationen erläutert. Ziel der vorliegenden Untersuchung ist die *Evaluation der Wirksamkeit* des FOSIS2-Programms im Hinblick auf die motivationalen Effekte bei den Schülern. Deshalb werden auch die Merkmale einer gelungenen *Wirksamkeitsevaluation* dargestellt.

#### 3.1.1 Definition von Evaluation

Allen Definitionen des Begriffs Evaluation ist gemein, dass sie grundsätzlich die Bewertung oder Beurteilung eines Sachverhalts beinhalten (Stockmann & Meyer, 2014). Entsprechend bezeichnet Helmke (2014) mit Evaluation die „Beurteilung des Wertes eines Produktes, Prozesses oder Programmes unter Zugrundelegung eines Gütemaßstabes“ (S. 272) und die Deutsche Gesellschaft für Evaluation (2008) definiert Evaluation ähnlich als „die systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes“ (S. 15). Der Gütemaßstab beziehungsweise die Kriterien, die für die Bewertung herangezogen werden, hängen vom jeweiligen Evaluationsgegenstand ab. Häufig steht der Nutzen für bestimmte Personen oder Gruppen im Vordergrund. *Wissenschaftliche Evaluationen* zeichnen sich durch bereits im Vorfeld festgelegte Kriterien und Standards aus, die mithil-

fe von Methoden der Sozialwissenschaften überprüft werden. Nach Stockmann und Meyer (2014) sind wissenschaftlich Evaluationen dadurch gekennzeichnet,

*„dass sie (1) auf einen klar definierten Gegenstand (z. B. politische Interventionsmaßnahmen, Projekte, Programme, Policies etc.) bezogen sind, (2) für die Informationsgenerierung objektivierende empirische Datenerhebungsmethoden eingesetzt [werden] und (3) die Bewertung anhand explizit auf den zu evaluierenden Sachverhalt und anhand präzise festgelegter und offengelegter Kriterien (4) mit Hilfe systematisch vergleichender Verfahren vorgenommen wird. Die Evaluation wird (5) in der Regel von dafür besonders befähigten Personen (Evaluatoren) durchgeführt (6) mit dem Ziel, auf den Evaluationsgegenstand bezogene Entscheidungen zu treffen“ (S. 74).*

Auch der Begriff der *Evaluationsforschung* wird üblicherweise nur auf solche Evaluationsstudien bezogen, die sich systematisch und korrekt empirischer Forschungsmethoden und -techniken bedienen (Bortz & Döring, 2006; Mittag & Hager, 2000; Rossi, Lipsey & Freeman, 2004; Wottawa & Thierau, 2003). Anders als die akademische Grundlagenforschung ist die Evaluationsforschung in der Regel anwendungsbezogen und zielt darauf ab, den praktischen Nutzen eines Evaluationsgegenstands zu ermitteln (Bortz & Döring, 2006). Zudem ist die Evaluationsforschung häufig durch einen konkreten Auftraggeber gekennzeichnet (Bortz & Döring, 2006; Stockmann, 2006; Stockmann & Meyer, 2014).

### **3.1.2 Funktionen von Evaluationen**

Evaluationen dienen der „Überprüfung von theoretischen Annahmen“ und sind „Entscheidungshilfe“ (Bauer, 2007), das heißt, sie liefern Argumente, die eine Entscheidungsvorbereitung, -begründung, -bewertung und -durchsetzung erleichtern (Wottawa & Thierau, 2003). Mithilfe dieser Argumente lassen sich je nach Evaluationsgegenstand und -ziel Aussagen treffen über die Wirksamkeit (Effektivität) z. B. einer Therapie und über deren Nachhaltigkeit, über die Kosten-Nutzen-Bilanz (Effizienz), die Qualität und Akzeptanz einer Maßnahme und auch darüber, ob überhaupt Bedarf für die Implementation einer solchen Intervention besteht (Gollwitzer & Jäger, 2009). Stockmann und Meyer (2014) schreiben Evaluationen übergreifend vier Funktionen zu, die *Erkenntnis-, Kontroll-, Entwicklungs- und die Legitimierungsfunktion*: Besteht das Ziel einer Evaluation darin, Informationen über die Akzeptanz der Intervention, die Bedürfnisse der Zielgruppe oder auch über die Rahmenbedingungen, die Umsetzung und die Wirkungen eines Programms zu sammeln, kommt ihr eine Erkenntnisfunktion zu. Sollen die Beteiligung und Einhaltung bestimmter mit der Intervention verbundener Verpflichtungen und Aufgaben überprüft

werden, wie z. B. bei einer Rechts- oder Wirtschaftsprüfung, so hat die Evaluation vor allem eine Kontrollfunktion inne. Liegt das Hauptinteresse in der Weiterentwicklung eines Programms, steht die Entwicklungsfunktion im Vordergrund. Eine zentrale Funktion von Evaluationen ist des Weiteren die Legitimation von Programmen und Maßnahmen, indem deren Effektivität nachgewiesen wird. Diese Funktion kann auch einen negativen Beigeschmack erhalten, wenn mithilfe von Evaluationen nachträglich politische Entscheidungen legitimiert werden sollen (Stockmann & Meyer, 2014).

### 3.1.3 Ziele von Evaluationen

Gollwitzer und Jäger (2009) heben drei Evaluationsziele hervor: *Zustandsevaluationen* dienen der Analyse von Merkmalen und Eigenschaften des zu evaluierenden Objekts. *Veränderungsevaluationen* zielen darauf ab, die Veränderungen eines Evaluationsgegenstands über die Zeit hinweg zu erfassen, und auch *Wirksamkeitsevaluationen* ermitteln Veränderungen, allerdings in der Regel solche, die in Folge einer Intervention entstanden sind.

Bezogen auf eine Interventionsmaßnahme beziehungsweise ein Programm können Evaluationen nach Stockmann und Meyer (2014) auf die *Programmentwicklung*, die *Implementation* eines Programms und schließlich auf das *Ergebnis* einer durchgeführten Maßnahme fokussieren:

Ist die Programmentwicklung, das heißt die Interventionsplanung und -konzeption Fokus der Evaluation, kann die Antizipation möglicher negativer Effekte Ziel sein, weshalb solche Untersuchungen auch *ex-ante Evaluationen* heißen (Stockmann & Meyer, 2014). Unter dem Fokus der Programmentwicklung sollen mithilfe von sogenannten *prognostischen* Evaluationen auch häufig bestimmte Zustände vorhergesagt werden (z. B. das Wahlergebnis bei politischen Wahlen). Die oben beschriebene Zustandsevaluation fällt folglich unter diesen Fokus. Auch die Prüfung des Implementationsbedarfs oder der Rahmenbedingungen und Voraussetzungen, die für die Implementation des Programms notwendig erscheinen, unterliegen dem Fokus der Programmentwicklung. Solche Evaluationen werden auch *prospektive Evaluation* oder *Inputevaluation* genannt (Gollwitzer & Jäger, 2009).

Dient eine Studie der Evaluation der *Implementation* einer Intervention, werden Informationen über den Verlauf und die Ergebnisse der Intervention beziehungsweise des Programms gesammelt. Ziele der Evaluation sind in diesem Fall die Kontrolle der Programmdurchführung und die Verbesserung der Durchführung der Intervention (Stockmann, 2006) beziehungsweise die Optimierung der Maßnahme oder des Programms selbst (Gollwitzer

& Jäger, 2009). Solchen Evaluationsstudien werden auch Begriffe wie *Prozess-* oder *Zwischenevaluation* (Gollwitzer & Jäger, 2009), *Implementationsevaluation* oder *-kontrolle*, *Programm Monitoring* (Westermann, 2002), *formative* (Gollwitzer & Jäger, 2009; Helmke, 2014; Köller, 2009; Mittag & Hager, 2000), *on-going* oder auch *aktiv-gestaltende Evaluation* zugeordnet (Stockmann, 2006; Stockmann & Meyer, 2014).

Steht das Ergebnis einer Intervention im Zentrum der evaluativen Untersuchungen, sollen in der Regel die Effekte der Maßnahme *ex-post* erfasst und bewertet werden (Stockmann, 2006). Dies entspricht der oben nach Gollwitzer und Jäger (2009) beschriebenen Wirksamkeitsevaluation, die auch *summative* (Helmke, 2014; Köller, 2009; Mittag & Hager, 2000; Westermann, 2002), *Ergebnis-*, *Produkt-* oder *Outputevaluation* (Westermann, 2002) genannt wird, wobei Mittag und Hager (2000) darauf hinweisen, dass auch Ergebnis-evaluationen formativ, also gestaltend angelegt sein können und die synonyme Verwendung der Begriffe *summative* und *Ergebnisevaluation* nicht immer angemessen ist.

Unterschieden werden zudem *globale* versus *analytische* Evaluationen. Globale Evaluationen verfolgen das Ziel, den Evaluationsgegenstand als Ganzes zu beurteilen. Im Gegensatz dazu werden bei analytischen Evaluationen einzelne Bestandteile des Evaluationsgegenstands bewertet (Köller, 2009; Westermann, 2002).

### **3.1.4 Rahmenbedingungen und Merkmale von Evaluationen**

Evaluationen unterscheiden sich in ihren Rahmenbedingungen: sie können im *Feld* oder im *Labor* stattfinden und werden auch als *geschlossen*, *offen*, *vertraulich* oder *öffentlich* bezeichnet. Die meisten Evaluationen laufen im realen Forschungsfeld ab, seltener sind Evaluationen im Labor, die experimentelle Maßnahmen bewerten. Bei geschlossenen Evaluationen ist die Fragestellung des Evaluationsprojekts im Vorhinein genau festgelegt, bei offenen Evaluationen hingegen ist gerade die Bestimmung der Fragestellung, der Methoden und der Hypothesen Ziel der Evaluation. Ergebnisse vertraulicher Evaluationen werden nur gegenüber dem Auftraggeber bekannt gegeben, eine öffentliche Evaluation ist dagegen allen Interessierten zugänglich. Manchmal wird das Begriffspaar „geschlossen“ versus „offen“ anstelle der Begriffe „vertraulich“ und „öffentlich“ verwendet (Gollwitzer & Jäger, 2009).

Insbesondere bei Evaluationen im Bildungskontext ist es nach Helmke (2014) außerdem wichtig, die Evaluationsebene zu beachten. So können die zu evaluierenden Interventionen, wie das FOSIS2-Programm, auf der *Mikroebene*, z. B. den Klassen, ansetzen oder

aber die Veränderung einer ganzen Schule (*Mesoebene*) oder eines Bildungssystems (*Makroebene*) zum Ziel haben (Burkard & Eikenbusch, 2007; Helmke, 2014).

Bei der Durchführung einer Evaluationsstudie werden in der Regel bestimmte Phasen durchlaufen. Nach Helmke (2014, S. 276) sind Evaluationen durch folgenden Ablauf gekennzeichnet: Planung der Evaluation (1), Bestimmung des Ziels (2) und der Zielgruppe (3), Konzeption der Durchführung (4), Datenerfassung (5), Auswertung und Interpretation der Daten (6), Anwendung der Ergebnisse (7) und Verbesserung des Programms (8). Evaluationsstudien werden auch hinsichtlich ihrer Durchführungshäufigkeit kategorisiert: Manche Evaluationen werden *einmalig* (Bauer, 2007) beziehungsweise *singulär* (Helmke, 2014) durchgeführt und andere finden *wiederholt* oder *kontinuierlich* (Bauer, 2007) beziehungsweise *zyklisch* (Helmke, 2014) statt.

An einer Evaluation sind darüber hinaus zahlreiche verschiedene Interessengruppen („Stakeholder“) beteiligt (z. B. Auftraggeber, Durchführende, Zielgruppe) (Stockmann & Meyer, 2014; Westermann, 2002). Im Rahmen von *parteilichen* Evaluationen werden z. B. vornehmlich die Interessen bestimmter Gruppen verfolgt. Bei sogenannten *ausbalancierten* Evaluationen hingegen sollen die Interessen aller an der Evaluation beteiligten Parteien gleichberechtigt berücksichtigt werden (Westermann, 2002). Unterschieden werden auch *interne* und *externe* Evaluationen. Interne Evaluationen werden von der Institution durchgeführt, in der auch das zu evaluierende Programm implementiert wird (Köller, 2009; Stockmann & Meyer, 2014). Eine spezifische Form der internen Evaluation ist die sogenannte Selbstevaluation, bei der die Evaluation durch Akteure erfolgt, die unmittelbar in die zu evaluierenden Prozesse involviert sind (Bauer, 2007; Stockmann & Meyer, 2014). Interne Evaluationen, die auch *In-house-Evaluationen* genannt werden, haben den Vorteil, dass die Evaluatoren meist Experten für den jeweiligen Evaluationsgegenstand sind und die hausinternen Abläufe der Institution gut kennen. Ein Nachteil ist dabei jedoch häufig die mangelnde Methodenkompetenz der Evaluatoren, die sich in einer geringeren Qualität der Evaluationsergebnisse niederschlagen kann. Zudem fehlt es den internen Evaluatoren oft an Unabhängigkeit, um die Programmeffekte frei von eigenen Interessen beurteilen zu können. Externe Evaluationen (auch *Fremdevaluationen* genannt; Westermann, 2002) werden von Personen durchgeführt, die keine Mitglieder der Organisation sind und somit in der Regel unabhängiger agieren können (Bauer, 2007; Stockmann & Meyer, 2014), wobei natürlich auch externe Evaluatoren parteilich sein können. Zumindest verfügen externe

Evaluatoren aber häufig über eine hohe Methodenkompetenz beziehungsweise sind speziell für die Durchführung von Evaluationen ausgebildet (Stockmann & Meyer, 2014).

Die Schilderung der hier genannten Rahmenbedingungen und Merkmale von Evaluationen hat nicht den Anspruch auf Vollständigkeit, sondern soll aufzeigen, dass verschiedene Unterscheidungsdimensionen existieren. Die Zuordnung einer Evaluation in die von einzelnen Autoren unterschiedlich gewählten Einteilungskategorien ist meist nicht eindeutig möglich und in der Praxis existieren häufig Mischformen (DeGEval – Deutsche Gesellschaft für Evaluation e. V., 2008).

### **3.1.5 Standards von Evaluationen**

Die Deutsche Gesellschaft für Evaluation (2008) fordert die Einhaltung folgender vier Standards für Evaluationen: Evaluationen sollten *nützlich*, *durchführbar*, *fair* und *genau* sein. Das Merkmal der Nützlichkeit weist darauf hin, dass eine Evaluation an den Bedürfnissen und dem Nutzen der Zielgruppe orientiert sein sollte. Hierzu müssen Ziel- und Betroffenenengruppen zunächst identifiziert und der Zweck der Evaluation beziehungsweise die damit verfolgten Ziele den Beteiligten transparent vermittelt werden. Zudem sollte die Evaluation rechtzeitig, das heißt für die Zielgruppe zu einem sinnvollen Zeitpunkt, an dem die Evaluationsergebnisse noch einen Einfluss auf künftige Entscheidungsprozesse nehmen können, umgesetzt werden. Die Durchführung erfolgt dabei möglichst durch methodisch kompetente Evaluatoren, die bei der Auswahl von erfassten Informationen und Methoden wissenschaftliche Standards und auch den Bedarf der Zielgruppe berücksichtigen. Die von der DeGEval (2008) empfohlenen Standards zur Durchführbarkeit sollen helfen, Evaluationen möglichst ökonomisch zu gestalten, sodass die Erfassung der Informationen für die betroffene Zielgruppe mit möglichst hoher Akzeptanz, geringfügigem Aufwand und einem angemessenen Verhältnis zwischen Kosten und Nutzen einhergeht. Darüber hinaus hat die DeGEval (2008) Fairnessstandards formuliert, nach denen Evaluationen die Vereinbarungen aller Stakeholder der Evaluation einhalten, die Rechte aller beteiligten Personen schützen und eine vollständige, unparteiische Durchführung der Evaluation sowie eine transparente Berichterstattung der Ergebnisse gewährleisten sollten. Schließlich werden im Rahmen der Genauigkeitsstandards der DeGEval (2008) eine exakte Beschreibung des Evaluationsgegenstands, -kontextes, -ziels und -vorgehens sowie die Offenlegung der Informationsquellen und natürlich eine valide und reliable Erfassung der Informationen gefordert. Mögliche Fehlerquellen sollten darüber hinaus identifiziert, die quantitativen und qualitati-



ven Ergebnisse genauestens analysiert und Schlussfolgerungen entsprechend begründet und dokumentiert werden.

### 3.1.6 Wirksamkeitsevaluation

Ziel der vorliegenden Arbeit ist die Evaluation der Wirksamkeit des FOSIS2-Programms. Deshalb sollen nachfolgend Merkmale dieser Evaluationsform näher beschrieben werden.

#### *Wirksamkeit, Wirkung und Wirkmodelle*

Gollwitzer und Jäger (2009) unterscheiden in Anlehnung an Hager und Hasselhorn (2000) die *Wirksamkeit* und die *Wirkung* einer Intervention: Mit Ersterer sind die mit einer Maßnahme intendierten Effekte gemeint, mit Letzterer die spezifischen *Wirkmechanismen*, die die beobachteten Effekte verursacht haben, wobei verschiedene Wirkungen beschrieben werden: Sogenannte *maßnahmenspezifische* Wirkungen bezeichnen die mit der Maßnahme erzielten und intendierten Wirkungen. *Neben- und Folgewirkungen* sind Wirkungen, die ebenfalls durch die Intervention verursacht wurden, jedoch nicht beabsichtigt waren. *Maßnahmenunspezifische* Wirkungen hingegen hängen nicht mit der Intervention selbst zusammen, sondern entstehen z. B. durch mit dem Programm verbundene Randbedingungen oder gar allein dadurch, dass eine Intervention stattgefunden hat. Auch *externe Wirkungen* lassen sich nicht auf die Durchführung der Intervention zurückführen, sondern auf andere extern gelagerte Störfaktoren (Gollwitzer & Jäger, 2009). Die Gesamtheit all dieser Wirkungen bezeichnen Hager und Hasselhorn (2000) als *Brutto-*, die maßnahmenspezifischen als *Nettowirkungen*.

Die jeweiligen Wirkmechanismen einer Maßnahme sollten in einem *Wirkmodell* dargestellt werden, das verdeutlicht, wie (Prozess) und wieso (Begründung) eine Maßnahme wirken soll und von welchen Randbedingungen ihre Wirkung abhängt (Gollwitzer & Jäger, 2009). Außerdem sollten alle im Modell spezifizierten unabhängigen und abhängigen Variablen genau operationalisiert werden. Die abhängigen Variablen sind die sogenannten Wirkkriterien. Wenn sich diese im Vergleich zum Ausgangszustand vor der Intervention entsprechend der Hypothesen verändert haben und die Veränderung nicht durch andere Einflussgrößen als die Intervention zustande gekommen ist, bestätigt dies die Wirksamkeit der Intervention. Erst eine exakte Operationalisierung der Wirkkriterien lässt die explizite Formulierung der Evaluationsziele zu. Bei der Operationalisierung der Kriterien und des Wirkmodells ist es wichtig, dass diese wirklich auf die Maßnahme abgestimmt wurden. Das heißt, Intervention und Wirkkriterien sollten *symmetrisch* sein. Symmetrie liegt vor,

wenn das Wirkkriterium genau das misst, was durch die Intervention beeinflusst werden soll, und weder zu spezifisch noch zu unspezifisch gewählt wurde (Bortz & Döring, 2006).

### ***Fragestellungen***

Im Rahmen einer Wirksamkeitsevaluation ist es einerseits möglich zu prüfen, ob die mit der Intervention intendierten Effekte eingetreten sind, und andererseits, ob die gefundenen Veränderungen tatsächlich durch die Maßnahme selbst verursacht wurden, es sich um maßnahmenspezifische Wirkungen handelt und ein kausaler Zusammenhang zwischen der Intervention und den abhängigen Variablen angenommen werden kann. Sollen beide Fragestellungen beantwortet werden, so liegen hierfür in der Regel gerichtete Hypothesen vor, die es zu testen gilt. Weitere Nebenfragen, die mit einer Wirksamkeitsevaluation geprüft werden können, betreffen die Nachhaltigkeit der Effekte („Halten die Effekte über einen gewissen Zeitraum an?“), den Transfer („Können die Interventionsteilnehmenden die erworbenen Fähigkeiten auch auf andere Situationen übertragen?“), die Erfassung der oben beschriebenen Neben- und Folgewirkungen und die Generalisierbarkeit der Ergebnisse („Sind die erreichten Effekte stichprobenspezifisch oder können sie auf andere Durchführungen der Interventionen übertragen werden?“) (Gollwitzer & Jäger, 2009, S. 80). Nach Hager (2008) kann im Rahmen von Wirksamkeitsevaluationen neben der Wirkung eines einzelnen Programms (*isolierte Wirksamkeitsanalyse*) auch die Wirkung eines Programms im Vergleich zu anderen Programmen (*vergleichende Wirksamkeitsanalyse*) evaluiert werden. Der isolierten Wirksamkeitsanalyse liegen einfache Wirksamkeitshypothesen (z. B. „Intervention A wirkt“) und der vergleichenden Wirksamkeitsevaluation Wirksamkeitsunterschiedshypothesen (z. B. „Intervention A wirkt stärker als Intervention B“) zugrunde.

### ***Design***

Das Design einer Wirksamkeitsevaluation sieht in der Regel eine Prä- und eine Postbefragung vor, zwischen denen die Interventionen umgesetzt werden (Hager, 2008; Köller, 2009). Häufig wird nach Beendigung von Intervention und Postbefragung auch noch eine dritte sogenannte Follow-up-Erhebung durchgeführt, um die Nachhaltigkeit der Programmeffekte zu untersuchen (Köller, 2009). Übliche Methoden, den Zustand der Studienteilnehmenden vor und nach der Intervention zu erfassen, sind Beobachtungen, Tests und Fragebogenerhebungen sowie Interviews (Øvretveit, 2002).

Inwieweit die erreichten Effekte auch kausal interpretiert, das heißt auf die alleinige Wirkung der durchgeführten Maßnahme zurückgeführt werden können, hängt von der *internen*

*Validität* ab (Gollwitzer & Jäger, 2009). Interne Validität ist gegeben, wenn die Veränderungen in den Wirkkriterien (den abhängigen Variablen) eindeutig durch die Wirkmechanismen der Interventionen (unabhängige Variablen) verursacht werden. Lässt sich nicht nur nachweisen, dass die erwarteten Effekte eingetreten sind, sondern auch, dass diese ohne Durchführung der Interventionen ausbleiben würden, kann von einer hohen internen Validität ausgegangen und ein kausaler Zusammenhang angenommen werden. Für diesen Nachweis wird neben der *Experimentalgruppe*, die das Treatment erhält, eine *Kontrollgruppe* benötigt, die zwar wie die Experimentalgruppe aus Personen der Zielpopulation zusammengesetzt ist und an der Prä- und der Post-Befragung teilnimmt, aber kein Treatment (*unbehandelte* Kontrollgruppe) beziehungsweise eine andere Maßnahme (*behandelte* Kontrollgruppe) erhält. Werden die erwarteten Veränderungen nur in der Experimentalgruppe, nicht aber in der Kontrollgruppe gefunden, lässt sich dies als ein Hinweis auf einen möglichen kausalen Zusammenhang zwischen Intervention und erreichten Effekten werten (Bortz & Döring, 2006; Gollwitzer & Jäger, 2009). Sollen im Sinne der isolierten Wirksamkeitsanalyse die Effekte eines einzelnen Programms ermittelt werden, empfiehlt Hager (2008) die Einführung einer behandelten Kontrollgruppe, die eine Maßnahme mit anderen Zielen, aber gleichen Randbedingungen wie die zu evaluierende Intervention erhält. Erst dadurch lässt sich ausschließen, dass die erreichten Effekte nur entstanden sind, weil überhaupt eine Intervention stattgefunden hat (Gollwitzer & Jäger, 2009), und die maßnahmen-spezifischen Nettowirkungen können ermittelt werden. Werden im Sinne der vergleichenden Wirksamkeitsanalyse mehrere Programme bezüglich ihrer Wirksamkeit verglichen, sollten diese Alternativinterventionen gleiche Programmziele verfolgen, können aber in ihren Randbedingungen in Abhängigkeit ihrer interventionsspezifischen Merkmale variieren (Hager, 2008).

Doch selbst wenn das Evaluationsdesign eine Kontrollgruppe beinhaltet, kann die interne Validität unterschiedlich stark ausgeprägt sein und durch verschiedene Störfaktoren verringert werden (Bortz & Döring, 2006; Gollwitzer & Jäger, 2009). Hager und Hasselhorn (2000) fassen unter dem Begriff der Störfaktoren alle Prozesse zusammen, die zu einer Schwächung der intendierten Programmwirkungen beitragen. Solche Störvariablen können z. B. externe Einflüsse, Reifungsprozesse, Testübungs-, instrumentelle, Regressions- (Chua, Tie & Don, 2011) und Selektionseffekte (Gollwitzer & Jäger, 2009) sowie fehlende Werte durch Drop-out-Prozesse sein (Köller, 2009). Im Zeitraum einer Depressionstherapie könnte z. B. ein lang anhaltender Wechsel von schlechtem zu sehr gutem Wetter stattgefunden haben. Es wäre denkbar, dass nicht die zu evaluierende Depressionstherapie,

sondern der Wetterwechsel zu einer Verbesserung der Depressionswerte geführt hat (externer Einfluss) (Chua et al., 2011). Die Untersuchungsteilnehmenden könnten sich auch unabhängig von der Intervention verändern und z. B. reifer werden (Köller, 2009). Durch die Teilnahme an der Prä-Befragung treten eventuell Übungseffekte, Erinnerungs- und Sensibilisierungseffekte auf, die die Post-Werte erhöhen, ohne dass die Intervention selbst an diesem Werteanstieg beteiligt sein muss. Auch durch mangelnde instrumentelle Reliabilität (instrumenteller Effekt) oder der Regression zur Mitte (Regressionseffekt) kann die interne Validität verringert werden (Chua et al., 2011). Ist die Teilnahme an einer Intervention freiwillig (Selbstselektion), kann es vorkommen, dass sich nur Personen mit bestimmten Eigenschaften an der Maßnahme beteiligen, so dass die gemessenen Effekte gar nicht auf die Maßnahme selbst, sondern möglicherweise auf die besonderen Eigenschaften der Teilnehmenden zurückzuführen sind (Gollwitzer & Jäger, 2009). Ein weiteres Problem pädagogisch-psychologischer Interventionsstudien ist darüber hinaus der häufig unangemessene methodische Umgang mit fehlenden Werten, wie sie z. B. durch eine Verweigerung an der weiteren Studienteilnahme durch einzelne Probanden oder auch durch Krankheit oder anderweitig begründete Teilnahmeausfälle zustande kommen können. Wenn möglich, sollten die fehlenden Werte analysiert, begründet und mittels der multiplen Imputation oder des Maximum-Likelihood-Verfahrens behandelt werden (Köller, 2009).

Es gibt verschiedene Möglichkeiten, die Störvariablen zu kontrollieren. Wichtigstes Verfahren ist die *Randomisierung*. Dabei werden die Personen nach dem Zufallsprinzip der Experimental- und Kontrollgruppe zugeteilt (Bortz & Döring, 2006). Durch Berücksichtigung des Zufalls werden die Störvariablen und damit auch ihr Einfluss über alle Gruppen gleich verteilt und somit „neutralisiert“ (Bortz & Döring, 2006, S. 54). Konnte randomisiert werden, liegt ein *experimentelles Design* vor (Bortz & Döring, 2006). Hager (2000) weist darauf hin, dass eine Randomisierung allein jedoch noch nicht automatisch zu einer vollständigen Kontrolle der Störvariablen führt, sondern wie oben beschrieben auch die Rahmenbedingungen der Vergleichsgruppen (z. B. unbehandelt versus behandelt) über die Kontrolle der Störvariablen mitentscheiden. In der Praxis der Evaluationsforschung ist eine Randomisierung aufgrund von ethischen Problemen oft sehr schwierig: Personen können nicht zu einer Teilnahme an einer Maßnahme gezwungen werden. Außerdem werden in der Evaluationsforschung meist Feldstudien durchgeführt, die nicht in künstlichen Laborsettings stattfinden. Hierbei können oft nur bestimmte Personengruppen erreicht werden, so dass praktisch gar nicht die Möglichkeit der Ziehung einer Zufallsstichprobe besteht. Solche sogenannten *quasi-experimentellen Designs*, in denen die Gruppenzuteilung nicht

nach dem Zufallsprinzip erfolgt (Bortz & Döring, 2006), sind typisch für Evaluationsstudien im Kontext Schule. Hier bilden in der Regel die Schulklassen bereits natürliche Gruppen, die sich nicht einfach unterteilen lassen, und insbesondere Lehrerfortbildungsmaßnahmen unterliegen den oben beschriebenen Selektionseffekten, sodass sich beispielsweise vermehrt Lehrkräfte an der Intervention beteiligen, die besonders motiviert und offen für Innovationen sind (Bauer, 2007). Solche Selektionsprozesse und andere mögliche Störfaktoren quasi-experimenteller Feldstudien sollten bei der Interpretation der Ergebnisse beachtet werden (Bortz & Döring, 2006). Bei der Durchführung von Evaluationen im Bildungskontext empfiehlt Helmke (2014) darüber hinaus, möglichst unterschiedliche qualitative und quantitative Methoden zu verwenden und die verschiedenen Perspektiven der Akteure in der zu untersuchenden Institution zu berücksichtigen.

### **3.1.7 Zusammenfassung: Evaluationsforschung**

Im Unterkapitel 3.1 wurden zentrale Merkmale der Evaluationsforschung beschrieben. Nachfolgend werden diejenigen Merkmale zusammenfassend dargestellt, die auch auf das FOSIS2-Projekt und die Untersuchungsanlage der vorliegenden Studie zutreffen.

Der Arbeit liegt das eingangs beschriebene Verständnis von wissenschaftlichen Evaluationen zugrunde. Nach Stockmann und Meyer (2014) können Evaluationsstudien als wissenschaftlich bezeichnet werden, wenn sie sich empirischer Forschungsmethoden bedienen und die Bewertung anhand a priori festgelegter Kriterien erfolgt. Zur Evaluation der Wirksamkeit des FOSIS2-Programms werden entsprechend dieser Forderungen empirische Methoden eingesetzt (Mehrebenenanalysen) und die Zielkriterien der vorliegenden Untersuchung (Selbstwirksamkeitserwartung, Prüfungsängstlichkeit, Stresserleben und Binnendifferenzierung) wurden in den vorhergehenden Kapiteln theoretisch hergeleitet und begründet. Hinsichtlich der beschriebenen Funktionen von Evaluationen wird hier vor allem die Erkenntnisfunktion verfolgt. Denn durch die Studie sollen Erkenntnisse über die FOSIS2-Programmeffekte gewonnen werden. In diesem Sinne ist auch schon das Ziel der Studie benannt, die Durchführung einer globalen (Köller, 2009), summativen (Helmke, 2014) Wirksamkeitsevaluation. Hinsichtlich der Rahmenbedingungen ist die vorliegende Felduntersuchung als eine singuläre (in dieser Form einmalig durchgeführte), geschlossene (die Fragestellungen sind im Vorhinein festgelegt) und öffentliche (die Ergebnisse werden öffentlich publiziert) Evaluation zu bezeichnen, die auf der Mikroebene (Schulklassen) ansetzt (Gollwitzer & Jäger, 2009; Helmke, 2014). Die Evaluation des FOSIS2-Programms erfolgt durch externe Evaluatoren, die keine Mitglieder der beteiligten Schulen sind

(Stockmann & Meyer, 2014). Da die Evaluatoren das FOSIS2-Programm jedoch auch selbst konzipiert und in die Schulen eingeführt haben, handelt es sich nicht um vollkommen unabhängige Evaluatoren.

Das Studiendesign der Untersuchung kann als ein quasi-experimentelles Kontrollgruppendesign charakterisiert werden, mithilfe dessen im Sinne einer isolierten Wirksamkeitsevaluation die Frage nach den Effekten des FOSIS2-Programms beantwortet werden soll (Hager, 2008). Eine behandelte Kontrollgruppe, wie sie von Hager (2008) gefordert wird, um die maßnahmenspezifischen Nettowirkungen zu ermitteln, konnte aus ökonomischen Gründen nicht umgesetzt werden.

### **3.2 Implementationsforschung**

In mehreren Studien konnte nachgewiesen werden, dass insbesondere bei komplexen, aus mehreren Modulen bestehenden Interventionsprogrammen die intendierten Programmwirkungen eng mit der Programmdurchführung, der *Implementation*, zusammenhängen (Durlak & DuPre, 2008; Kalafat et al., 2007). Für die Programmwirksamkeit ist entscheidend, wie und unter welchen Bedingungen die einzelnen Programmkomponenten in das Feld implementiert wurden (Gollwitzer & Jäger, 2009). Eine fehlgeschlagene Implementation kann die Wirksamkeit einer Intervention schwächen (Beelmann & Karing, 2014; Durlak, 2013; Durlak & DuPre, 2008). Rossi, Lipsey und Freeman (2004) beschreiben drei Wege der inadäquaten Implementation: Die Intervention wird gar nicht oder nur sehr wenig umgesetzt (1). Es wird nicht das richtige Programm implementiert (2). Das Programm wird nicht standardisiert durchgeführt beziehungsweise die Umsetzung unterscheidet sich innerhalb der Zielpopulation (3). Durch eine solche inadäquate Implementation kann eine Intervention fälschlicherweise als unwirksam bewertet werden (Domitrovich & Greenberg, 2000; Rossi et al., 2004). Daher ist das frühzeitige und regelmäßige Monitoring der Implementation zu empfehlen (Durlak, 2013; Durlak & DuPre, 2008; Gollwitzer & Jäger, 2009; Weissberg, Kumpfer & Seligman, 2003). Insbesondere wenn Interventionen durch vermittelnde Personen (Mediatoren) umgesetzt werden, ist eine Evaluation der Implementation unerlässlich, da Ausmaß und Genauigkeit der Umsetzung zwischen den Mediatoren variieren können (Durlak, 2013; Durlak & DuPre, 2008; Petermann, 2014b; Scheirer et al., 1995). Studien belegen darüber hinaus, dass sich die Implementation von Innovationen,

speziell im Kontext Schule als schwierig erweist (Payne, 2009; Payne et al., 2006)<sup>21</sup>. In einem Review, in dem die Wirksamkeit von 213 in der Schule implementierten Interventionen zum sozialen und emotionalen Lernen untersucht wurde, fanden sich bei Schülern, die an gut umgesetzten Programmen teilnahmen, deutlich bessere Leistungen und eine stärkere Reduktion von negativen Emotionen in der Schule als bei Schülern, die an schlecht umgesetzten Programmen beteiligt waren (Durlak, Weissberg, Dymnicki, Taylor & Schellinger, 2011). Gerade bei schulbasierten Präventionsprogrammen sollte daher nicht nur die Wirksamkeit, sondern auch die Implementation des Programms evaluiert werden (Durlak & DuPre, 2008; Gollwitzer & Jäger, 2009; Weissberg et al., 2003). Trotz dieses Wissens um die Bedeutung der Implementation wird diese nur selten systematisch untersucht und gerade in Bezug auf die Schul- und Unterrichtsforschung ist hier ein Forschungsdesiderat zu konstatieren (Fries & Souvignier, 2009; Gräsel & Parchmann, 2004; Hasselhorn et al., 2014; Owens et al., 2014). Die Forderung nach einer stärkeren Beforschung von Implementationsprozessen wird in der Arbeit aufgegriffen und es wird explorativ der Zusammenhang zwischen der Umsetzung des FOSIS2-Programms und Programmeffekten untersucht. Nachfolgend werden deshalb zusammenfassend wesentliche Erkenntnisse der Implementationsforschung beschrieben und nach einer Begriffsdefinition mögliche Kennwerte zur Beurteilung der Implementation, Implementationsstrategien, Methoden zur Erfassung der Implementation und Rahmenbedingungen für eine erfolgreiche Umsetzung erläutert.

### 3.2.1 Definition von Implementation

Fixsen et al. (2005) definieren Implementation in ihrer Literaturanalyse zur Implementationsforschung wie folgt: „Implementation is defined as a specified set of activities designed to put into practice an activity or program of known dimensions“ (Fixsen et al., 2005, S. 5). Auch die „Umsetzung von Neuerungen beziehungsweise Innovationen in die Praxis“ wird mit dem Begriff der Implementation gekennzeichnet (Hasselhorn et al., 2014, S. 141). Euler und Sloane (1998) verstehen unter Implementation noch allgemeiner die „Umsetzung von wissenschaftlichen Erkenntnissen in die gesellschaftliche Praxis“ (S. 312). Dieser Definition entsprechend lässt sich die *Implementationsforschung* als eine in-

---

<sup>21</sup> Ein großes Hindernis für erfolgreiche Umsetzung schulischer Fortbildungsprogramme stellt beispielsweise die zeitliche Belastung der teilnehmenden Lehrkräfte dar. Je länger die Fortbildung andauert, desto höher ist die zeitliche Beanspruchung und desto geringer fallen die Akzeptanz gegenüber der Fortbildung und desto schlechter die Umsetzung der Fortbildungsinhalte aus (Fries & Souvignier, 2009). „Bei der Implementation neuer Programme muss daher das sensible Gleichgewicht zwischen zeitlicher Belastung und zur Qualitätssicherung notwendiger Investition von Zeit gewahrt werden“ (Fries & Souvignier, 2009, S. 425).

terdisziplinär angelegte Wissenschaft fassen, die sich mit dem Transfer von Forschungsbefunden in die Praxis beschäftigt und versucht, die in vielen Disziplinen beschriebene „Forschungs-Praxis-Lücke“ zu überwinden (Petermann, 2014a, S. 121).

### 3.2.2 Implementationskennwerte

In der Implementationsforschung wurden bereits zahlreiche Kennwerte identifiziert, die zur Beurteilung der Umsetzung einer Maßnahme in die Praxis geeignet sind. Bisher hat sich jedoch keine einheitliche Begriffstaxonomie durchgesetzt und unterschiedliche Autoren kennzeichnen ähnliche Prozesse und Kennwerte der Implementation mit verschiedenen Begrifflichkeiten (Beelmann & Karing, 2014; Petermann, 2014b). Am häufigsten werden die Implementationsquantität und -qualität als Kennwerte der Implementation benannt und untersucht. Unter *Quantität* beziehungsweise *Umsetzungshäufigkeit*, *Dosis* oder *Bandbreite* (*,quantity, dosage, intervention strength, level of implementation‘*) wird das Ausmaß beziehungsweise der Grad der Implementation des Programms verstanden. Durch die *Qualität* (*,quality‘*) wird erfasst, wie gut einzelne Programmaspekte umgesetzt wurden (Durlak & DuPre, 2008; Petermann, 2014b). Weitere von Durlak und DuPre (2008) in einem Review zum Einfluss der Implementation auf die Programmwirksamkeit ermittelte Kennwerte sind die *Genauigkeit* (*,fidelity‘*), die *Akzeptanz der Programmteilnehmer* (*,participant responsiveness‘*), die *Einzigartigkeit des Programms* (*,program differentiation, uniqueness‘*), das *Monitoring*, die *Reichweite* (*,program reach‘*) und die *Anpassung* (*,adaption, modification, reinvention‘*). Mit der Genauigkeit wird ermittelt, inwieweit die Implementation des Programms der Programmbeschreibung beziehungsweise dem Manual entspricht. Die Akzeptanz der Programmteilnehmer misst, ob das Programm die Interessen beziehungsweise die Aufmerksamkeit der Programmteilnehmer weckt und aufrechterhält. Mithilfe der Einzigartigkeit des Programms wird dargestellt, ob sich das Programm hinsichtlich seiner theoretischen Konzeption und der praktischen Umsetzung von anderen Programmen unterscheidet. Unter Monitoring verstehen die Autoren die Überwachung von Kontroll- und Vergleichsbedingungen beziehungsweise -gruppen. Die Reichweite beschreibt das Ausmaß der Programmbeteiligung. Schließlich verdeutlicht die Anpassung die Veränderungen des Originalprogramms während der Durchführung (Durlak & DuPre, 2008). Auch Petermann (2014b) klassifiziert in einem Überblicksartikel zum Stand der Implementationsforschung in Anlehnung an Grimshaw et al. (2006) und Michie et al. (2005; 2009) die Genauigkeit beziehungsweise Wiedergabetreue sowie die Akzeptanz als



Kennwerte der Implementation.<sup>22</sup> Darüber hinaus benennt er die *Übernahme* (*,adoption‘*), *Angemessenheit* (*,appropriateness‘*), *Machbarkeit* (*,feasibility‘*), *Implementationskosten* (*,cost‘*), *Durchdringung* (*,penetration‘*) und die *Nachhaltigkeit* (*,sustainability‘*) als weitere Implementationskennwerte. Unter Übernahme wird die eigentliche Entscheidung für eine Programmumsetzung verstanden (Petermann, 2014b). Hierunter fällt auch das von Durlak und DuPre (2008) beschriebene Ausmaß der Umsetzung. Die Angemessenheit charakterisiert die Passung zwischen einer spezifischen Maßnahme und dem jeweiligen Anwendungskontext. Mit der Angemessenheit verwandt, aber doch inhaltlich von ihr zu trennen, ist die Machbarkeit. Diese verdeutlicht, wie leicht sich eine Maßnahme in einem bestimmten Setting umsetzen lässt. Während die Angemessenheit also angibt, ob eine Intervention für bestimmte Anliegen der Zielinstitution sinnvoll erscheint, beschreibt die Machbarkeit, inwieweit die Maßnahme praktisch realisierbar ist. Im Rahmen von Kosten-Nutzen-Studien werden die Implementierungskosten untersucht. Das Merkmal der Durchdringung ähnelt der oben beschriebenen Reichweite und bezeichnet das Ausmaß der Integration einer Intervention in die Zielinstitution. Schließlich wird mit dem Kennwert der Nachhaltigkeit überprüft, ob eine Maßnahme auch langfristig umgesetzt und aufrechterhalten wird (Petermann, 2014b). Neben den geschilderten Kennwerten ist die sogenannte *Tiefe* (*,depth‘*) der Umsetzung ein weiteres Kriterium, das für die Beurteilung der Implementation herangezogen werden kann. Die Tiefe beschreibt, inwieweit eine Neuerung nicht nur die Praktiken der Anwender verändert hat, sondern auch Einfluss auf deren Einstellungen, Werte und Überzeugungen genommen hat (Hasselhorn et al., 2014).

In den 59 von Durlak und DuPre (2008) in ihrem Review analysierten Studien wird am häufigsten die Implementationsgenauigkeit (37 Studien) untersucht, gefolgt von der Implementationsquantität, also dem Grad der Umsetzung (29 Studien). Es zeigt sich, dass eine perfekte Implementation nicht zu erwarten ist. In der Regel liegt das Ausmaß der Umsetzung zwischen 60 und 80 %.

### 3.2.3 Implementationsstrategien

Bezogen auf die Unterrichts- beziehungsweise die Lehr-Lernforschung unterscheiden Gräsel und Parchmann (2004) drei verschiedene Implementationsstrategien, die sogenannten *Top-Down*-, *Bottom-up*- und die *symbiotischen* Strategien: Bei Ersterer wird die zu imple-

---

<sup>22</sup> Petermann (2014b) bezeichnet die Kennwerte zur Beurteilung des Implementationsprozesses als Implementation outcomes.

mentierende Innovation von externen Personen auf der Basis von wissenschaftlichen Erkenntnissen entwickelt. Im Gegensatz dazu werden beim Bottom-Up-Vorgehen die Interventionen intern von den Anwendern selbst entwickelt. Mit symbiotischen Strategien werden Maßnahmen bezeichnet, die von Wissenschaftlern, Lehrkräften und anderen Akteuren des Bildungskontextes gemeinsam konzipiert und umgesetzt werden. Da das Bottom-up-Vorgehen in der Regel Einzelfälle von Schulentwicklungsmaßnahmen beinhaltet, lassen sich hier keine allgemeinen Merkmale hinsichtlich der Implementation ableiten, weshalb Gräsel und Parchmann (2004) vornehmlich die Top-Down- und die symbiotischen Strategien beschreiben:

Der Erfolg der Umsetzung von Innovationen nach dem Top-Down-Prinzip wird in der Regel anhand des oben beschriebenen Kennwerts der Genauigkeit gemessen.<sup>23</sup> Solche Interventionsbemühungen scheitern häufig an der mangelnden Berücksichtigung des Erfahrungswissens der Lehrkräfte. Beurteilen diese das Programm als irrelevant oder wenig nützlich, wirkt sich dies negativ auf die Programmumsetzung aus. Die klassische Top-Down-Strategie wird daher zunehmend aufgebrochen und die Lehrkräfte stärker in die Programmkonzeption einbezogen, so dass Programmadaptionen,<sup>24</sup> die die Besonderheiten der schulischen Praxis berücksichtigen, möglich werden.

Bei der symbiotischen Implementationsstrategie wird eine Innovation von Anfang an gemeinsam von allen beteiligten Akteuren und Interessengruppen auf der Basis eines als relevant erachteten Problems entwickelt. Zentrales Merkmal der symbiotischen Programmumsetzung ist die Kooperation dieser Akteure, die häufig über einen längeren Zeitraum anhält und eine kontinuierliche Weiterentwicklung der Innovation ermöglicht, gestützt durch wissenschaftliche Erkenntnis, Erfahrungswissen und Erprobung des Programms im Schulalltag. Kooperationsstrukturen sowie die Verringerung beziehungsweise Verbesserung des Ausgangsproblems und die Verbreitung des Programms sind Kennwerte, anhand derer der Erfolg der symbiotischen Implementation einer Maßnahme gemessen wird (Gräsel & Parchmann, 2004).

---

<sup>23</sup> Snyder, Bolin und Zuwalt (1992) sprechen deshalb auch von der sogenannten *Fidelity-Perspektive*.

<sup>24</sup> Snyder, Bolin und Zuwalt (1992) nennen diese erweiterte Implementationsstrategie auch *„mutual adaption“*.

### 3.2.4 Methoden zur Erfassung der Implementation

Die Implementation wird nach Durlak und DuPre (2008) in der Regel auf zwei Wegen untersucht: Entweder es werden Gruppen von Programmumsetzern (Trainern, Mediatoren) erzeugt, die sich in ihrer Programmimplementation unterscheiden. Oder die Implementation wird kontinuierlich erfasst. Dabei wird ein multimethodales Vorgehen, das quantitative und qualitative Methoden beinhaltet, empfohlen (Kalafat et al., 2007). Für die Erfassung einer Programmumsetzung kommen Aufzeichnungen beziehungsweise Dokumentationen, Beobachtungen und Selbstberichte in Frage. Wird beispielsweise die Häufigkeit der Interaktionen zwischen Programmadministratoren und Programmumsetzern dokumentiert, kann dies Aufschluss über eine gelungene Implementation geben. Die Daten aus solchen Dokumentationen sind in der Regel objektiv und reliabel (Petermann, 2014b). Auch Beobachtungsdaten sind eine valide Quelle zur Erfassung der Implementation. Die Beobachtung ist jedoch sehr ressourcenaufwändig. Wesentlich einfacher und praktikabler ist die Messung der Umsetzung mittels Selbstberichten (Gollwitzer & Jäger, 2009; Petermann, 2014b), z. B. durch die Befragung der Trainer und Programmumsetzer. Hierbei ist jedoch durch motivationale und kognitive Verzerrungen mit Einbußen in der Validität zu rechnen. Auch die eigentliche Zielgruppe der Intervention kann befragt werden, was ebenfalls messfehleranfällig, jedoch reliabler ist als eine Befragung der Trainer (Gollwitzer & Jäger, 2009). Durlak und DuPre (2008) empfehlen auch die Durchführung möglichst periodisch, kontinuierlich eingesetzter Kurzbefragungen (sogenannter *„spot checks“*) zur Implementationsevaluation.

### 3.2.5 Rahmenbedingungen und Merkmale für eine erfolgreiche Implementation

Die Implementation eines Programms basiert im besten Fall auf einer spezifischen Beschreibung der einzelnen Programmkomponenten. Gegebenenfalls mündet diese in ein Programmmanual, das Interventionsinhalte und -abläufe standardisiert (Gollwitzer & Jäger, 2009). Eine verständliche, anschauliche und anwendungsfreundliche Programmbeschreibung erhöht die Umsetzungstreue und die Qualität der Implementation (Beelmann & Karing, 2014). Je genauer die Durchführung der Maßnahme den Manual-Vorgaben entspricht, desto erfolgreicher beziehungsweise wirksamer ist sie auch (Gollwitzer & Jäger, 2009; Petermann, 2014b). Wandersman et al. (2008) betonen, dass die Identifikation und Beschreibung von sogenannten Kernelementen eines Programms zudem zu einer höheren Akzeptanz und Umsetzungsbereitschaft seitens der Programmanwender führen. Bei der Programmbeschreibung sollte grundsätzlich deutlich werden, welche Maßnahmen mög-

lichst konzeptgetreu implementiert werden sollten und wo ggf. Abweichungen erlaubt sind (Petermann, 2014b). Gewisse Anpassungen und Adaptionen der Intervention an die realen Bedingungen der jeweiligen Zielinstitution haben sich als förderlich erwiesen, solange die Kernelemente eines Programms entsprechend der Manualvorgaben umgesetzt werden (Beelmann & Karing, 2014). Es existieren verschiedene Rahmenmodelle, die versuchen, den Prozess einer erfolgreichen Programmimplementation zu beschreiben (Durlak, 2013). Meyers, Durlak und Wandersman (2012) haben in einem Literaturüberblick zu den bereits existenten Modellen übergreifende Schritte identifiziert, die eine hohe Umsetzungsqualität gewährleisten. Diese stellen sie zusammenfassend in ihrem Phasenmodell „Quality Implementation Framework (QIF)“ vor. Hiernach ist in der Phase vor der eigentlichen Programmumsetzung zunächst zu klären, inwieweit das Zielprogramm in den zu untersuchenden Kontext passt, ob Programmadaptionen für den Kontext nötig sind und welche Erwartungen beziehungsweise Befürchtungen die Programmanwender haben. Zudem ist eine genaue Planung der Umsetzung erforderlich, sodass die Anwender in der Phase der realen Implementation des Programms Unterstützung erhalten und die Umsetzungsaktivitäten supervidiert und evaluiert werden können (Meyers et al., 2012).

In ihrem Review zur Implementationsforschung identifizierten Durlak und DuPre (2008) verschiedene Merkmale, die zu einer erfolgreichen Implementation beitragen. Diese lassen sich folgenden Bereichen zuordnen:

- *Merkmale der Gemeinde (,community factors‘)*
- *Merkmale der Programmdurchführenden (,provider characteristics‘)*
- *Merkmale des Programms (,innovation characteristics‘)*
- *Merkmale der Organisation (,organizational capacity‘)*
- *Merkmale der Unterstützung (,training and technical assistance‘)*

Entscheidende, die Implementation beeinflussende Merkmale der Gemeinde sind beispielsweise politische Gelegenheitsstrukturen und Strategien (Durlak & DuPre, 2008). So hängen die für eine Präventionsmaßnahme bereitgestellten Ressourcen auch von politischen Entscheidungen ab (Beelmann & Karing, 2014). Für die Programmdurchführenden gilt: Personen, die den Sinn, den Bedarf und die Notwendigkeit für die Implementation sehen, die glauben, dass die Innovation die erwünschten Effekte bringt, die hoch selbstwirksam sind und die die nötigen Fähigkeiten zur Umsetzung des Programms besitzen, implementieren das Programm in der Regel in einem hohen Ausmaß und mit Genauigkeit (Beelmann & Karing, 2014; Durlak & DuPre, 2008). Im Bereich der Innovationsmerkmale

sind Flexibilität und Kompatibilität mit dem Kontext als für die Implementation bedeutsame Faktoren zu nennen. Hinsichtlich der Organisationsmerkmale haben sich solche Institutionen als implementationsförderlich erwiesen, die ein Klima der Innovationsoffenheit kultivieren und eine effektive Organisationsführung sowie mindestens einen Programmbefürworter aufweisen, der möglichst bei allen Mitarbeitern anerkannt ist. Zudem ist ein Klima der Kooperation, in dem das Prinzip des sogenannten ‚shared-decision-making‘ praktiziert wird, eine gute Voraussetzung für eine erfolgreiche Implementation. Merkmale zur Unterstützung sind die für das Programm und die Umsetzung zur Verfügung stehende Zeit, vorhandene administrative und finanzielle Ressourcen, die Mitarbeiterausstattung sowie die Motivation und das Commitment der Mitarbeiter. Unter idealen Bedingungen interagieren diese fünf Bereiche und führen zu einer effektiven Implementation des Programms (Durlak & DuPre, 2008). In anderen Reviews zur Implementation werden ähnliche Faktoren als Voraussetzung für eine erfolgreiche Implementation benannt: die Ausstattung, ein positives Arbeitsklima, geteiltes Entscheiden, Koordination mit anderen Organisationen, konkret formulierte Aufgaben für die Umsetzung, eine effektive Führung, die Existenz von Programmbefürwortern innerhalb der Organisation, administrative Unterstützung und seitens der Programmdurchführenden die nötigen Fähigkeiten zur Umsetzung (Fixsen et al., 2005; Greenhalgh, Robert, MacFarlane, Bate & Kyriakidou, 2004; Stith et al., 2006). Die Fähigkeiten zur Umsetzung sollten im Rahmen von gut durchdachten Schulungs-, Weiterbildungs- und Supervisionsangeboten trainiert werden (Beelmann & Karing, 2014).

Ähnliche Rahmenbedingungen haben sich auch für eine erfolgreiche Implementation von Innovationen im Kontext Schule als bedeutsam erwiesen. So tragen nach Fries und Souvignier (2009) folgende Maßnahmen zur nachhaltigen Implementation von Interventionen in der Schule bei: Die Lehrkräfte sollten bei der Auswahl einer Fortbildung einbezogen und die zeitliche Belastung durch die Teilnahme an der Fortbildung so gering wie möglich gehalten werden. Darüber hinaus sollten im Rahmen einer Einführungsveranstaltung zunächst die der jeweiligen Intervention zugrunde liegenden theoretischen Konzepte und deren jeweilige Bedeutung für die pädagogische Praxis vermittelt werden. Einfach verständliches und leicht handhabbares Material, das die Interventionsmaßnahme beschreibt, verbessert ebenfalls das Umsetzungsverhalten und auch Wiederholungssitzungen erhöhen die nachhaltige Implementation. Wichtig sind zudem die Möglichkeit der Reflexion von Schwierigkeiten bei der Umsetzung und eine wissenschaftliche Begleitung (Fries & Souvignier, 2009). Payne und Kollegen (2006; 2009) haben die Implementation von Präventi-

onsprogrammen an 544 Schulen in den USA untersucht und ebenfalls geprüft, welche Rahmenbedingungen die Programmimplementation erhöhen. Zusätzlich zu den bereits berichteten Merkmalen scheinen die Integrationsmöglichkeit des Programms in den Schulalltag, das Ausmaß der Unterstützung durch die Schulleitung und allgemein die Intensität des Trainings und der Supervision die Implementation zu beeinflussen (Payne & Eckert, 2009; Payne et al., 2006). Nach Sun, Shek und Siu (2008) sind für eine erfolgreiche Programmimplementation in der Schule außerdem folgende Faktoren entscheidend: eine gute räumliche Ausstattung; möglichst kein autoritäres, sondern ein positives Arbeitsklima, in dem offen kommuniziert und zusammen gearbeitet werden kann; Leidenschaft und Überzeugung seitens der beteiligten Lehrkräfte, Jugendliche fördern zu wollen; eine einladende schulische Atmosphäre („school of care and love“), in der sich die Schüler als Teil der Schule fühlen („sense of belonging“) (S. 1068) und von ihren Mitschülern wertgeschätzt werden; eine hohe Identifikation der Schule mit dem Programm, das heißt übereinstimmende Grundeinstellungen und Werte zwischen Schule und Programm; ein großes Engagement und eine hohe Innovationsbereitschaft der beteiligten Lehrkräfte und Schüler sowie eine ausreichende personale und materielle Ausstattung (Sun et al., 2008). Darüber hinaus beeinflussen auch die Einstellungen und subjektiven Theorien der beteiligten Lehrkräfte gegenüber dem Programm die Implementation. Akzeptieren die Lehrkräfte die Intervention und erleben sie diese als sinnvoll und umsetzbar, so wird die Programmumsetzung verstärkt (Gräsel & Parchmann, 2004). Auch eine enge Kooperation der beteiligten Lehrkräfte sowie deren gegenseitige Unterstützung fördern die Implementationsqualität. Teamentwicklungsmaßnahmen zur Verbesserung der Kooperation in einem Kollegium sollten im Rahmen der Umsetzung von Präventionsprogrammen in der Schule daher explizit angeregt werden (Beelmann & Karing, 2014).

### **3.2.6 Zusammenfassung: Implementationsforschung**

Die Wirksamkeit eines Programms hängt eng mit der Qualität der Programmumsetzung, der Implementation zusammen (Durlak, 2013). In der vorliegenden Untersuchung wird deshalb die Wirksamkeit des FOSIS2-Programms auch unter Berücksichtigung der Implementation untersucht. Der Arbeit liegt dabei das von Fixsen et al. (2005) beschriebene Implementationsverständnis zugrunde, wonach mit Implementation alle Aktivitäten bezeichnet werden, die eine Intervention in die Praxis umsetzen. Es existieren verschiedene Kennwerte, anhand derer die Implementation eines Programms charakterisiert und bewertet werden kann (vgl. z. B. Durlak & DuPre, 2008; Petermann, 2014b). In dieser Arbeit

wird das Ausmaß der Umsetzung untersucht. Hinsichtlich der von Gräsel und Parchmann (2004) beschriebenen Implementationsstrategien kann das FOSIS2-Programm der Top-Down-Strategie zugeordnet werden, wobei die beteiligten Lehrkräfte die eigentliche Umsetzung der einzelnen FOSIS2-Maßnahmen auch mitgestalten konnten. Hinsichtlich der Methoden zur Erfassung und Evaluation der Implementation empfehlen Durlak und DuPre (2008) die Durchführung möglichst periodisch, kontinuierlich eingesetzter Kurzbefragungen, sogenannter spot checks. Diese Empfehlung wurde im FOSIS2-Projekt aufgegriffen, sodass im Laufe der zwei Projektjahre alle drei Monate kurze Checkbefragungen zur Erfassung der Umsetzung des Programms durchgeführt wurden. In der Implementationsforschung konnten verschiedene Rahmenbedingungen und Merkmale identifiziert werden, die zu einer erfolgreichen Implementation beitragen. Im FOSIS2-Projekt wurde versucht, diese zu berücksichtigen und somit die Programmumsetzung zu verbessern. Entsprechend der Empfehlungen von Fries und Souvignier (2009) wurden die Lehrkräfte in die Gestaltung der Fortbildungen eingebunden und die zeitliche Belastung sollte durch möglichst kurze Workshops, die direkt im Anschluss an den Unterricht stattfanden, gering gehalten werden. Zu jedem FOSIS2-Modul wurden Handreichungen ausgegeben, die sowohl die theoretischen Konzepte als auch die praktischen Förderstrategien zusammenfassend und leicht verständlich darstellen. Im Rahmen von Beratungen und Supervision durch die wissenschaftliche Begleitung war zudem die Möglichkeit der Reflexion von Schwierigkeiten bei der Umsetzung gegeben. Durch einen frühen Kontakt mit der Schulleitung sollte deren Unterstützung bei der Programmumsetzung, wie von Payne und Kollegen (2006; 2009) gefordert, sichergestellt werden.

Bevor weitere Merkmale der Umsetzung des FOSIS2-Programms ausführlicher in Kapitel 6 geschildert werden, werden in den nachfolgenden Kapiteln 4 und 5 zunächst das Gesamtprojekt FOSIS2 und die Fragestellungen der vorliegenden Untersuchung beschrieben.

## **4 Das FOSIS2-Projekt**

Bei dem Projekt „Förderung von Schutzfaktoren in der Schule 2 (FOSIS2)“ handelt es sich um eine quasi-experimentelle Längsschnittstudie, die am Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie der Humboldt-Universität zu Berlin in Kooperation mit der Unfallkasse Berlin – unterstützt durch den Berliner Senat – konzipiert und durchgeführt wurde. Ziel des Projekts war es, bei Schülern der siebten und achten Jahrgänge dreier Berliner Oberschulen verglichen mit einer Kontrollgruppe günstigere Entwicklungen von motivationalen und sozialen Ressourcen beziehungsweise Risikofaktoren zu erreichen. Im Verlauf zweier Schuljahre wurden hierzu Lehrkräfte als Mediatoren im Rahmen von sechs Modulworkshops geschult, durch ein modifiziertes Unterrichtsverhalten und die Einführung von Förderstrategien in ihren regulären Fachunterricht die motivationalen und sozialen Ressourcen und Risikofaktoren ihrer Schüler zu beeinflussen. In der vorliegenden Arbeit soll die Wirksamkeit des FOSIS2-Programms im Hinblick auf die motivationalen Effekte evaluiert werden. Bei der Darstellung der praktischen Förderstrategien werden deshalb vor allem die gewünschten motivationalen Auswirkungen und der Zusammenhang zu den in dieser Arbeit untersuchten Zielkriterien erläutert. Zuvor werden jedoch zunächst der Bezug des FOSIS2-Projekts zu Vorgängerprojekten und der zeitliche Verlauf des Projekts dargestellt.

### **4.1 Bezug und Abgrenzung zu Vorgängerprojekten**

Die Konzeption und Evaluation von Interventionen zur Ressourcenstärkung bei Schülern haben eine langjährige Forschungstradition am Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie der Humboldt-Universität zu Berlin. Interventionsprojekte wie „SIGIS – Sicher und gesund in der Schule“ (Jerusalem, 2005a), „FOSS – Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht“ (Jerusalem et al. 2007) und „FOSIS1 – Förderung von Schutzfaktoren in der Schule 1“ (Jerusalem, 2008) bilden die Vorläufer für die Entwicklung des FOSIS2-Programms. Die Ergebnisse der Evaluationen dieser Vorgängerprojekte belegen die Wirksamkeit von Lehrerfortbildungsprogrammen im Hinblick auf eine Stärkung motivationaler und sozialer Ressourcen von Schülern (z. B. Drössler, 2009; Drössler, Röder & Jerusalem, 2007; Röder, 2008; Röder, Drössler & Jerusalem, 2010). Darüber hinaus konnten im Rahmen der Evaluationen jedoch auch Schwachstellen und daraus abgeleitete Verbesserungsmöglichkeiten für eine erfolgreiche Programmdurchführung identifiziert werden. Besonders große Schwierigkeiten ergaben sich



bei der Implementation der Förderstrategien in den Unterricht und bei der Erfassung beziehungsweise der Evaluation des Implementationsgeschehens. So wurde in den Vorgängerprojekten eine teilweise unzureichende und mangelhafte Implementation der Fördermaßnahmen konstatiert. Diese Schwierigkeiten können auf eine zu geringe Bereitschaft zu einer verbindlichen Projektteilnahme seitens der beteiligten Schulen, auf eine zu komprimierte und kompakte Vermittlung der Förderstrategien sowie auf fehlende Unterstützung bei der Umsetzung der Interventionsinhalte in den schulischen Alltag und einer damit verbundenen zu geringen Teilnahme- und Durchhalte-Motivation der beteiligten Lehrkräfte zurückgeführt werden. Bei der Konzeption des FOSIS2-Projekts wurden die in den Vorgängerprojekten identifizierten Schwachstellen berücksichtigt. Durch Veränderungen in den Teilnahmebedingungen, in der zeitlichen und organisatorischen Struktur der Interventionsinhalte und durch die Einführung eines umfassenden Unterstützungs- und Beratungsangebots während der Programm-Umsetzungsphasen sowie der intensiven und häufigen Erfassung des Implementationsgeschehens sollte eine Wiederholung der Probleme aus den Vorgängerprojekten vermieden werden.

### ***Veränderte Teilnahmebedingungen***

Zur Förderung einer verbindlichen Teilnahme über den gesamten Projektzeitraum wurde bei der Akquise potentieller Projekt-Schulen bereits in ersten Gesprächen mit der Schulleitung die Einhaltung bestimmter Rahmenbedingungen als Teilnahmevoraussetzung ausgehandelt: Dazu gehörte die freiwillige Beteiligung der Lehrkräfte. Lehrer der zu Beginn des Projekts siebten Jahrgänge sollten sich selbstbestimmt, ohne äußeren Druck, aber in Kenntnis aller Rahmenbedingungen des Projekts aktiv zu einer Teilnahme entscheiden können. Die Schulleitung wurde gebeten, die Projektbeteiligung der Lehrkräfte zu unterstützen, ohne Kollegen zu einer Projektteilnahme zu verpflichten. Dadurch sollten die Teilnahme- und Durchhalte-Motivation der beteiligten Lehrkräfte erhöht, die Umsetzungsrate der Interventionsinhalte gesteigert und die Qualität der Implementation verbessert werden. Teil der Unterstützung seitens der Schulleitung und weitere Voraussetzung für eine schulische Projektbeteiligung war zudem die Ermöglichung von klassenübergreifenden Lehrerteams sowie die Einführung entsprechender Zeitfenster für eine solche Teamarbeit. In den Vorab-Gesprächen mit der Schulleitung wurde außerdem darauf hingewiesen, dass eine Entscheidung für eine Projektteilnahme seitens der Schule zu einer vollständigen Beteiligung über die gesamte Projektdauer von zwei Schuljahren verpflichtet. Die vereinbarten Teilnahmebedingungen wurden in entsprechenden Kooperationsverträgen mit den

Projektpartnern (die jeweilige Schule, die Humboldt-Universität zu Berlin, die Unfallkasse Berlin) schriftlich festgehalten und von allen Verantwortlichen im Sinne eines verbindlichen Vertrags unterzeichnet.

### ***Veränderte zeitliche und organisatorische Struktur der Interventionsinhalte***

In den Vorgängerprojekten wurde den beteiligten Lehrkräften in umfassenden Fortbildungsveranstaltungen eine Vielzahl verschiedener Förderstrategien inklusive praxisbezogener Anwendungsaufgaben vermittelt. Dadurch hatten die Lehrkräfte die Möglichkeit, theoretische und inhaltliche Bezüge zwischen den einzelnen Fördermaßnahmen herzustellen. Durch die komprimierte Darstellung der Inhalte und die Fülle an Förderstrategien war es den Lehrkräften jedoch kaum möglich, ein tieferes Verständnis von den Interventionsinhalten zu gewinnen und sich intensiv mit einzelnen Strategien auseinanderzusetzen. So fanden dann vor allem eher einfache und den Lehrkräften vertraute Maßnahmen Eingang in den Unterricht und komplexe, noch unbekannte Förderinhalte blieben unberücksichtigt. Im FOSIS2-Projekt sollten diese Probleme durch eine grundlegend veränderte Struktur der Fortbildung verhindert werden. Dazu wurden die Fördermaßnahmen nach inhaltlichen Gesichtspunkten in sechs kompakte, inhaltlich in sich kohärente Förderbereiche unterteilt. Diese wurden den Lehrkräften im Rahmen von sechs Fortbildungsworkshops vermittelt, die im Abstand von circa drei Monaten stattfanden. Dadurch hatten die Projektlehrer jeweils drei Monate Zeit, sich mit einzelnen Förderbereichen intensiv zu befassen und die gelernten Strategien sukzessive in den Unterricht zu integrieren. Vor dem eigentlichen Programmbeginn fanden in den Projektschulen Informationsveranstaltungen statt, in denen die sechs Förderbereiche bereits überblicksartig vorgestellt wurden, sodass den Lehrkräften auch im FOSIS2-Projekt weiterhin die Möglichkeit erhalten blieb, theoretische und inhaltliche Bezüge zwischen den einzelnen Förderstrategien zu erkennen.

### ***Einführung eines umfassenden Unterstützungs- und Beratungsangebots während der Programmumsetzungsphasen***

Die in den Vorgängerprojekten im Hinblick auf Quantität und Qualität als ungenügend identifizierte Implementation der Programminhalte sollte im FOSIS2-Projekt nicht nur durch eine verbindlichere Regelung der Teilnahme und durch eine veränderte Struktur der Fortbildung, sondern auch durch die Einführung eines umfassenden Unterstützungs- und Beratungsangebots während der Programmumsetzungsphasen verbessert werden. Der Einsatz neuer, unbekannter Unterrichtsstrategien erfordert einen erhöhten Aufwand, eine Ab-

kehr von alten, bekannten Routinen, eine Neuorganisation des Unterrichts und ein höheres Maß an Selbstreflexion und aktiver Selbstregulation im eigenen unterrichtlichen Handeln. Diese motivationalen und volitionalen Hürden der Lehrkräfte sind mögliche Gründe für eine mangelnde Umsetzung von Förderstrategien. Zudem sind die positiven Auswirkungen der Interventionsinhalte seitens der motivationalen und sozialen Ressourcen der Schüler oft erst nach einer langfristigen, regelmäßigen und qualitativ hochwertigen Implementation der Interventionsinhalte in den Unterricht zu erwarten, während unmittelbare Wirkungen oft ausbleiben. Die am Projekt beteiligten Lehrkräfte sehen darin oft die Erfolglosigkeit ihres Bemühens bestätigt und kehren zu den alten Unterrichtsrouinen zurück. Im FOSIS2-Programm sollte deshalb ein umfassendes Unterstützungs- und Beratungsangebot implementiert werden, das den Projektlehrern bei der Überwindung der genannten Schwierigkeiten und Hürden helfen sollte. Dazu wurden insbesondere Maßnahmen zur sozialen Unterstützung konzipiert, die in dieser Form in den Vorgängerprojekten noch nicht umgesetzt werden konnten: Jede am Projekt teilnehmende Interventionsschule sollte dauerhaft von einem Mitarbeitenden der Humboldt-Universität zu Berlin wissenschaftlich beraten und begleitet werden. In Einzel-, Gruppen-, E-Mail- und Telefon-Beratungen konnten die Projektlehrer mit ihrem jeweiligen „Schul-Betreuer“ Schwierigkeiten bei der Umsetzung der Interventionsinhalte besprechen und beseitigen. Darüber hinaus wurden Unterrichtshospitationen sowie die unterstützende Teilnahme der wissenschaftlichen Begleitung bei Elternabenden und Schulinformationstagen angeboten. Zudem wurden auch kollegiale Unterstützungsangebote durch die Förderung der Bildung von klassenübergreifenden, zum Teil fachhomogenen Projektteams und entsprechende Teamtreffen angeregt. Die Unterstützungs- und Beratungsangebote wurden zeitlich und im Format flexibel den Bedürfnissen der jeweiligen Projektschulen angepasst. Zur Erleichterung der Kommunikation zwischen wissenschaftlicher Begleitung und beteiligten Schulen sowie der Schulen untereinander wurde außerdem eine Homepage des FOSIS2-Projekts entwickelt, auf der den Schulen der Zugriff zu Projektinformationen sowie der Austausch der von den Projektlehrern erarbeiteten Arbeitsmaterialien ermöglicht wurde.

### ***Intensive und häufige Erfassung des Implementationsgeschehens***

Nicht nur die eigentliche Umsetzung der Interventionsinhalte in den Unterricht, sondern auch die Erfassung dieses Implementationsgeschehens war in den Vorgängerprojekten erschwert. Das Beantworten von Fragebögen zur Implementation sowie die Dokumentation des eigenen Umsetzungsverhaltens im Rahmen von Protokollen oder sogenannten Log-

büchern einzeln oder auch in Projektteams erwiesen sich für die beteiligten Lehrkräfte als ein erheblicher zeitlicher Zusatzaufwand. Die Rücklaufquoten waren gering. Im FOSIS2-Programm wurde versucht, diesen Erfahrungen Rechnung zu tragen, indem möglichst kurze Instrumente zur Erfassung der Implementation eingesetzt und nicht nur die Lehrkräfte, sondern auch die Schüler zum Umsetzungsgeschehen befragt wurden. Standardisierte Befragungen sollten vor allem die Quantität der Implementation erfassen und qualitative Interviews dienten der Erfassung von Schwierigkeiten bei der Umsetzung. Die Lehrkräfte wurden per E-mail angeschrieben und online standardisiert befragt, sodass der Beantwortungsaufwand im Vergleich zu den Befragungen der Vorgängerprojekte reduziert und gleichzeitig die Häufigkeit der Erfassung erhöht werden konnte. Auf diesem Weg wurde versucht, das Implementationsgeschehen differenzierter, umfassender und lückenloser als in den Vorgängerprojekten zu erfassen. Eine ausführliche Beschreibung der Instrumente zur Erfassung der Implementation, die als Datengrundlage für die vorliegende Untersuchung herangezogen werden, findet sich unter Abschnitt 6.4.2.

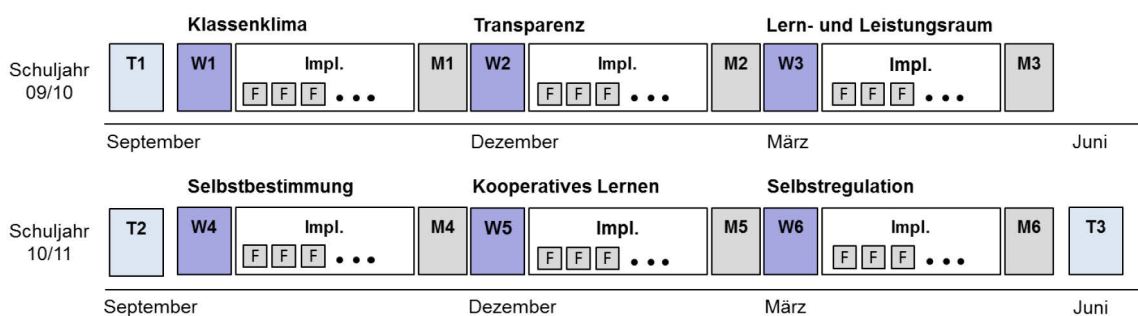
## **4.2 Zeitlicher Verlauf des Projekts**

Die Vorbereitungen des FOSIS2-Projekts begannen bereits zu Beginn des Jahres 2009. Bis zum Sommer dieses Jahres wurden die Workshops und die Instrumente zur Erfassung des Implementationsgeschehens und zur Evaluation des Projekts konzipiert sowie die teilnehmenden Schulen rekrutiert. Dabei erfolgte die Entwicklung und Auswahl der für die Motivation relevanten Förderstrategien und der Skalen für die Erhebungsinstrumente auf der Basis der unter Kapitel 2 geschilderten Theorien, theoretischen Konzepte und empirischen Erkenntnisse sowie auf der Basis von Erkenntnissen aus den Vorgängerprojekten. Als Interventionsschulen konnten eine Berliner Real- und zwei Berliner Gesamtschulen für die Teilnahme am Projekt gewonnen werden (eine ausführliche Beschreibung der Rekrutierung der Schulen findet sich unter 6.1.1). Eine weitere Real- und zwei Gesamtschulen wurden als Kontrollschulen rekrutiert, sodass sich insgesamt sechs Berliner Oberschulen mit 22 siebten und im zweiten Projektjahr achten Klassen an dem Projekt beteiligten. Die Kontrollschulen erhielten keine Intervention und beteiligten sich mit ihren Schülern und Lehrern der siebten Klassen nur an einzelnen Befragungen des FOSIS2-Projekts.

Nachdem kurz vor Ende des Schuljahres 2008/2009 in jeder Interventionsschule eine Informationsveranstaltung stattgefunden hatte, in der die Lehrkräfte über das FOSIS2-Projekt und die damit verbundenen Maßnahmen informiert wurden, begann die eigentliche Umsetzung des über zwei Schuljahre andauernden Projekts zu Beginn des neuen Schuljahres im

September 2009. Eine Übersicht über den zeitlichen Verlauf und die wesentlichen Elemente des FOSIS2-Projekts gibt Abbildung 2. Zur Evaluation des FOSIS2-Programms wurden vor Beginn der Intervention und nach dem ersten sowie nach dem zweiten Schuljahr standardisierte Befragungen mit Schülern und Lehrern der siebten und im zweiten Projektjahr dann achten Klassen der Interventions- und Kontrollschulen durchgeführt (in Abbildung 2 = T1 bis T3). Der Schülerfragebogen („paper-pencil“) umfasst z. B. Skalen zu Selbstwirksamkeitserwartungen, zur Prüfungsängstlichkeit und zum schulbezogenen Stresserleben (siehe auch Abschnitt 6.4.1). Der Lehrerfragebogen („online“) beinhaltet z. B. Skalen zu spezifischen Selbstwirksamkeitserwartungen, zu Aspekten des Unterrichtsverhaltens und Teilnahmemotiven sowie zum projektspezifischen Erfolgserleben. Zwischen den ersten beiden Befragungen (T1 und T2) wurden im ersten Projektjahr mit den Lehrkräften der Interventionsschulen drei Modulworkshops zu den Themen Klassenklima, Transparenz, Lern- und Leistungsraum durchgeführt (in Abbildung 2 = W1 bis W3). Im zweiten Projektjahr – zwischen der zweiten und dritten Befragung (T2 und T3) – fanden in jeder Interventionsschule drei weitere Modulworkshops zu den Themen Selbstbestimmung, Kooperatives Lernen und Selbstregulation statt (in Abbildung 2 = W4 bis W6). In den Workshops wurden neben theoretischen Inputs zu den einzelnen Themenbereichen die Umsetzung konkreter Förderstrategien in Lehrer-Teams gemeinsam geplant und erste Schritte für den eigenen Unterricht vorbereitet und geübt. Im Anschluss an jeden Workshop hatten die Lehrkräfte jeweils etwa drei Monate Zeit, die erarbeiteten Förderstrategien in ihren Fachunterricht einzuführen (in Abbildung 2 = Impl.). In den Implementationsphasen sollten die Lehrkräfte darüber hinaus die bereits erlernten Strategien aus vorangegangenen Workshops fortführen. Zur Erleichterung der Implementation erhielten die Lehrkräfte Handreichungen in Form eines Manuals. Zudem fanden die in Abschnitt 4.1 beschriebenen Einzel-, Gruppen-, Vor-Ort-, E-mail- und Telefon-Beratungen durch die FOSIS2-Projektmitarbeiter statt, in denen Umsetzungsschwierigkeiten der teilnehmenden Lehrkräfte besprochen wurden. Ergänzend konnten die Lehrkräfte das Angebot von Unterrichtshospitationen wahrnehmen und wurden durch Projektmitarbeiter auch bei der Informierung über das Projekt bei Elternabenden und Schulinformationstagen unterstützt. Die Teilnahme an den Unterstützungs- und Beratungsangeboten war freiwillig – die Formate wurden an die jeweiligen Bedürfnisse der Schulen angepasst. Zur Verbesserung der Kommunikation zwischen den Projektmitarbeitern und den beteiligten Schulen, sowie den teilnehmenden Lehrkräften untereinander existierte außerdem eine Webseite, auf der wichtige Informationen zum Projekt festgehalten und im Projekt erstellte Arbeitsmaterialien ausgetauscht werden konnten.

Im Kapitel 3.2 wurde bereits auf die Notwendigkeit hingewiesen, insbesondere bei schulbasierten Präventionsprogrammen nicht nur die Wirksamkeit, sondern auch die Implementation des Programms zu evaluieren (Durlak & DuPre, 2008; Gollwitzer & Jäger, 2009; Weissberg et al., 2003). Im FOSIS2-Projekt kamen verschiedenste Maßnahmen zur Erfassung der Implementation zum Einsatz: Mittels kurzer standardisierter wöchentlicher Feedbacks (online) sollten regelmäßige Informationen zu den laufenden Projektaktivitäten der Interventionslehrer erhoben werden (in Abbildung 2 = F). Zudem wurden die an der Intervention beteiligten Lehrer und Schüler am Ende jeder Implementationsphase in Form von standardisierten Befragungen und Interviews um sogenannte modulspezifische Rückmeldungen gebeten (in Abbildung 2 = M1 bis M6). Der standardisierte Schülerfragebogen (paper-pencil) zur modulspezifischen Rückmeldung erfasst das von den Schülern wahrgenommene Implementationsverhalten der Lehrer (eine ausführliche Beschreibung erfolgt in Kapitel 6.4.2). Der Lehrerfragebogen (online) beinhaltet Items zum Implementationsverhalten bezüglich Einsatzhäufigkeit, Einsatzbreite, auftretender Schwierigkeiten und der erlebten Unterstützung durch Kolleginnen und Kollegen, die Schulleitung und die wissenschaftliche Begleitung. Die Interviews fanden mit wechselnden Lehrern und Schülergruppen von 2-3 Teilnehmern statt. Erfragt wurden dabei die beiderseitigen Erfahrungen mit den Förderstrategien (z. B. Schwierigkeiten bei der Umsetzung und vermutete Effekte).



**Abbildung 2: Zeitlicher Verlauf des Projektes**

*Anmerkungen:* T = Prä-, Zwischen- und Postbefragung (mit Schülern und Lehrern), W = Workshop, Impl. = Implementationsphase, M = Modulspezifische Rückmeldungen zur Implementation (standardisierte Befragungen und qualitative Interviews mit Schülern und Lehrern); F = wöchentliches Feedback zur Implementation (nur mit Lehrern)

### 4.3 Beschreibung der Module

Die sechs Module des FOSIS2-Programms zu den Themen Klassenklima, Transparenz, Lern- und Leistungsraum, Selbstbestimmung, Kooperatives Lernen und Selbstregulation sind bewusst in der beschriebenen Reihenfolge eingeführt worden. Da im Land Berlin nach der sechsten Klasse der Schulwechsel auf die weiterführende Schule stattfindet, waren die

teilnehmenden siebten Klassen zu Projektbeginn neu zusammengesetzt worden. Zur Förderung eines „Wir-Gefühls“ und einer Gruppenkohäsion im Klassenverband wurde das Modul zum Klassenklima deshalb zu Beginn des Projekts implementiert. Die darauffolgenden Module sind nach der Komplexität ihrer Inhalte und ihrem Schwierigkeitsgrad in Bezug auf die Umsetzung gestaffelt. In den Modulen zur Transparenz und zum Lern- und Leistungsraum werden sehr konkrete Förderstrategien behandelt, die für die Lehrkräfte in ihrer Rolle als Vermittler ebenso gut handhabbar sein sollen wie für die Schüler als Empfänger. Die Module des zweiten Projektjahres fordern stärker als die ersten drei Module eine Internalisierung der Prinzipien der Förderstrategien und nehmen an Schwierigkeit bei der Umsetzung zu. So stellt das Modul zur Selbstregulation die höchsten Anforderungen an die Lehrkräfte, da hier keine einzelne Förderstrategie zur Anwendung kommt, sondern ein umfassendes handlungsregulatorisches Modell zur Ableitung praktischer Anknüpfungspunkte im Unterricht genutzt werden soll. Nachfolgend werden die Inhalte der einzelnen Module näher beschrieben.

#### **4.3.1 Klassenklima**

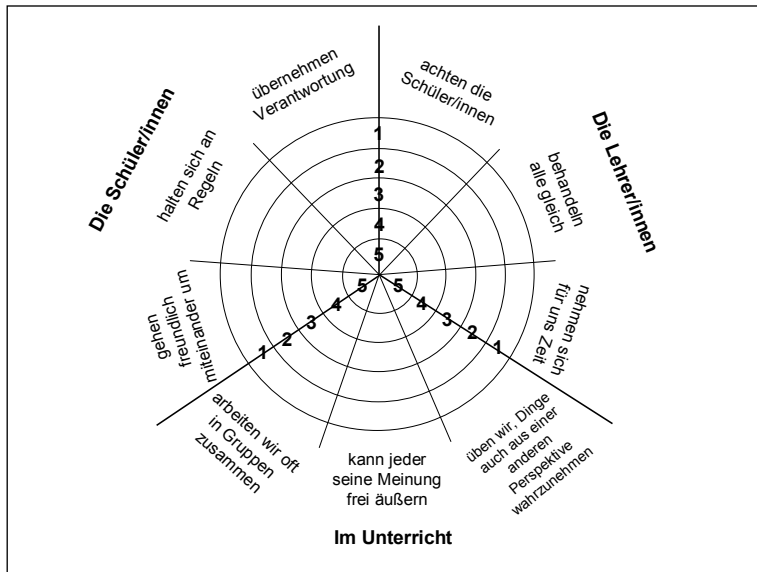
„[Das] *Klima* lässt sich als die von den Betroffenen wahrgenommene Konfiguration bedeutsamer Merkmale innerhalb der jeweiligen schulischen Umwelt beschreiben“ (Eder, 2010, S. 649). Die Definition verdeutlicht, dass es sich bei der Wahrnehmung des Klimas um ein subjektives Erleben handelt, das von verschiedenen Merkmalen abhängt. Diese lassen sich nach Saldern und Littig (1987) drei Bereichen zuordnen: der Schüler-Lehrer-Beziehung, der Schüler-Schüler-Beziehung und den Merkmalen des Unterrichts. Faktoren im Bereich der Schüler-Lehrer-Beziehung sind beispielsweise die Lehrerfürsorglichkeit, der Führungsstil der Lehrperson oder auch die Bevorzugung beziehungsweise die Benachteiligung von einzelnen Schülern. Hilfsbereitschaft, Konflikte oder auch die vorherrschende Konkurrenz zwischen den Schülern einer Klasse sind Merkmale der Schüler-Schüler-Beziehung. Beispiele für Merkmale des Unterrichts sind der Leistungsdruck, Mitbestimmungsmöglichkeiten oder auch die Möglichkeit zur Verantwortungsübernahme im Unterricht. In vielen Untersuchungen konnte die Bedeutsamkeit eines positiven Klassenklimas für das Erleben und Verhalten von Kindern und Jugendlichen sowie für eine gelingende unterrichtliche Interaktion nachgewiesen werden (vgl. Fraser & Walberg, 1991; Jerusalem & Schwarzer, 1991; König, 2009; Satow, 2002; Zumhasch, 2009). So zeigen sich positive Auswirkungen eines guten Klimas auf die Schulleistung, auf die Einstellungen gegenüber der Schule, auf Engagement und Lernverhalten sowie auf Selbstwert und Selbstkonzept

(Eder, 2010). Im Hinblick auf die in dieser Arbeit untersuchten Variablen belegen empirische Befunde außerdem, dass eine positive Veränderung in der kollektiven Wahrnehmung der Zufriedenheit mit dem Unterricht mit einer deutlichen Verbesserung der schulischen Selbstwirksamkeitserwartung einhergeht (Kolbe et al., 1998). Auch Satow (1999) weist auf den positiven Zusammenhang zwischen der schulischen Selbstwirksamkeitserwartung und einem sogenannten *Mastery-Klima* hin. Als *Mastery-Klima* bezeichnet Satow ein „Klima, das sich (a) durch individualisierte Lehrer-Schüler-Beziehungen, (b) durch supportive Schüler-Schüler-Beziehungen sowie (c) durch Unterrichtszufriedenheit und Autonomie auszeichnet“ (S. 69). Auch das schulische Belastungserleben verringert sich durch ein positives Klima. So nehmen in einem guten Klima Prüfungsängstlichkeit und schulbezogener Stress ab (Eder, 2010). In diesem Sinne sollte das erste Klima-Modul des FOSIS2-Programms nicht nur zu einer Verbesserung der sozialen Kompetenzen der Schüler und zur Stärkung des „Wir-Gefühls“ sowie der Gruppenkohäsion in den neu zusammengesetzten siebten Klassen beitragen, sondern bei den Interventionsschülern verglichen mit der Kontrollgruppe auch die schulische Selbstwirksamkeitserwartung sowie Prüfungsängstlichkeit und Stresserleben positiv beeinflussen.

Der Workshop lief konkret folgendermaßen ab: Zunächst wurden die von Saldern und Littig (1987) eingeführten Merkmalsbereiche des Klassenklimas und die geschilderten Auswirkungen und Korrelate eines positiven Klimas dargestellt. Bevor Maßnahmen zur Förderung des Klassenklimas eingesetzt werden, sollte eine Diagnose des aktuellen Klimas einer Klasse erstellt werden. Eine solche Klimadiagnose kann auf verschiedene Arten durchgeführt werden. So existieren beispielsweise Fragebögen oder Stimmungsskalen, die das Schul- oder auch das Klassenklima erfragen. Eine Übersicht über solche Instrumente hat die Projektgruppe „Feedbackmaterialien“ um Haider (2013) verfasst. Eine sehr praktische und plakative Methode, die im Rahmen des Modulworkshops Klassenklima eingeführt wurde, ist die Klimadiagnose mittels einer Zielscheibe. Diese ist wie eine Dartscheibe aufgebaut, wobei die Kreissegmente für unterschiedliche Merkmale des Klimas stehen. Die Merkmale können von der Lehrkraft im Vorfeld eingetragen oder gemeinsam mit den Schülern erarbeitet werden. Jeder Schüler erhält dann eine Kopie der Zielscheibe, wo er für die einzelnen Merkmale auf einer fünfstufigen Skala von „trifft nicht zu“ (1) bis „trifft zu“ (5) eine Beurteilung vornimmt. Die Voten der einzelnen Schüler können dann eingesammelt und auf eine große ausgehängte Zielscheibe übertragen werden, sodass sich ein Klassenstimmungsbild abzeichnet und deutlich wird, in welchen Merkmalen die Schüler einer Klasse überwiegend Defizite und somit noch einen Verbesserungsbedarf konstatieren. Eine



alternative Möglichkeit des Ausfüllens ist, eine große Zielscheibe von jedem Schüler zur Wahrung der Anonymität verdeckt hinter der Tafel ausfüllen zu lassen. Dieses Vorgehen führt aber möglicherweise zu Verzerrungen, da sich die Schüler mit ihren Voten gegenseitig beeinflussen können. Abbildung 3 veranschaulicht eine Zielscheibe grafisch. Im Workshop wurden die Lehrkräfte gebeten, diese Zielscheibe beispielhaft für eine Klasse auszufüllen und dann eine Zielscheibe mit eigenständig entwickelten Merkmalen zu versehen. Diese Zielscheibe sollte dann in den Klassen eingesetzt werden.



**Abbildung 3: Zielscheibe zur Klimadiagnose**

Auf der Basis der Ergebnisse der Klimadiagnose waren die Lehrkräfte aufgefordert, zur Verbesserung der als defizitär diagnostizierten Merkmale gemeinsam mit ihren Schülern Klassenregeln zu erstellen. Dabei sollte es sich nicht um Disziplinar- und Ordnungsregeln handeln, sondern um soziale Regeln, die auf die Gestaltung des Miteinanders in der Klasse abzielen. Im Workshop wurden Prüfpunkte erarbeitet, mithilfe derer „gut formulierte“ Regeln von „schlechten“ Regeln unterschieden werden können. Diese Prüfpunkte lauteten:

- Ist die Regel konkret formuliert, sodass die Schüler genau wissen, was von ihnen erwartet wird?
- Ist das gewünschte Verhalten unter willentlicher Kontrolle der Schüler?
- Ist das gewünschte Verhalten überprüfbar?
- Besteht Konsens über die Einhaltung und den Sinn der Regeln?
- Ist die Regel positiv formuliert?

Gut formulierte soziale Regeln beschreiben Verhalten konkret und eindeutig und beziehen sich auf willentlich kontrollierbare und überprüfbare Handlungen. Eine gelungene Regel

sollte außerdem für alle einen ersichtlichen Sinn ergeben und zum leichteren Verständnis sowie zur Vermeidung von Trotz- und Reaktanzreaktionen positiv formuliert sein (Werkstatt, 2001). Mit den Lehrkräften wurden außerdem Maßnahmen zur Umsetzung und zur Überprüfung der Einhaltung der Regeln erarbeitet. Beispielsweise wurde der Einsatz von wechselnden „Regelwächtern“ vorgeschlagen, die in den einzelnen Klassen jeweils für eine Woche auf die Einhaltung der Regeln achtgaben und auf Regelbrüche hinweisen sollten. Oder es konnte ein spezielles Zeitfenster („Fair Play“) von z. B. zehn Minuten in den Unterricht eingeführt werden, in dem alle gemeinsam explizit auf die Einhaltung der Regeln achten sollten. Zum Umgang mit Regelbrüchen wurden Verhaltensverträge sowie Punktesysteme mit entsprechenden Klassenbelohnungen (z. B. Ausflüge) erarbeitet und ein vierstufiges Intervenieren bei Regelbrüchen besprochen. Hiernach enthält die erste Stufe einen nonverbalen Hinweis auf den Regelbruch und die zweite eine verbale Rückmeldung, in der auf den Regelbruch hingewiesen wird. In der dritten Stufe wird der Schüler dann aufgefordert, sich zu entscheiden, ob er die Regel wieder einhalten oder andernfalls die vorher vereinbarte Konsequenz bei Regelbruch annehmen möchte. Sollte er sich für die Regeleinhaltung entscheiden und die Regel zu einem späteren Zeitpunkt wieder brechen, so würde er dann eine obligatorische Auszeit nehmen müssen.

Eine weitere Möglichkeit der Klimaverbesserung, die im Rahmen des Workshops vorgeschlagen wurde, stellt die Verantwortungsübernahme von Schülern für Schüler beziehungsweise für die Klasse dar. Übernehmen Schüler Verantwortung, z. B. in Form von Lernpatenschaften, ergibt sich die Gelegenheit, soziale Kompetenzen zu entwickeln, gegenseitige Unterstützung zu erfahren, Erfolge zu erleben und sich sozial in die Klasse eingebunden zu fühlen. Möglichkeiten zur Verantwortungsübernahme sind auch die schülergeleitete Vorbereitung und Organisation von Klassenfahrten oder anderen Klassen-Ereignissen, die Leitung von AGs oder Projektgruppen, Schülermoderation im Unterricht und vieles mehr.

#### **4.3.2 Transparenz**

Nach Jacobs (1997) werden mit dem Begriff Angst umgangssprachlich auch Unsicherheit und Ungewissheit assoziiert. Speziell in Prüfungssituationen kann Ungewissheit darüber bestehen, was in welcher Weise abgeprüft wird, wie eine effektive Vorbereitung gestaltet werden kann und nach welchen Kriterien bewertet wird. Um diese Unklarheiten zu beseitigen, stattdessen *Transparenz* und Klarheit zu steigern und dadurch auch die Angst vor Prüfungen zu reduzieren, wurde in der Arbeitsgruppe um Strittmatter das Transparenzpapier

entwickelt (siehe z. B. Strittmatter, 1997), das im FOSIS2-Workshop zum Thema Transparenz eingeführt wurde. Die Maßnahme wurde bereits ausführlich in Kapitel 2.3.6 vorgestellt. Transparenzpapiere sollen in der Schule rechtzeitig vor Prüfungen ausgeteilt werden und über die Anforderungen, sprich die Prüfungsthemen, die Vorbereitungsmöglichkeiten und Aufgabentypen sowie über die Bewertungskriterien, also die Zuordnungen von Punkten zu Aufgaben, die Notenvergabe und über die Konsequenzen der Leistungsbewertung, informieren. Aufgrund der teilweisen Überlappung der Konstrukte Prüfungsangst und Stresseinschätzung (Pekrun & Götz, 2006) lässt sich vermuten, dass der Einsatz von Transparenzpapieren nicht nur zu einer Reduktion des Angst-, sondern auch des Stresserlebens beiträgt. Darüber hinaus sollen Transparenzpapiere auch die eigenen Kompetenzerwartungen steigern. Denn Transparenzpapiere unterstützen Schüler darin, eine komplexe und schwierige Aufgabe, wie die Vorbereitung auf eine Klassenarbeit, in Nahziele zu untergliedern und Schritt für Schritt zu bewältigen. Die Untergliederung einer Aufgabe in Teilziele erhöht die Wahrscheinlichkeit für die Erfahrung eigener Erfolge, wodurch die schulische Selbstwirksamkeitserwartung von Schülern gesteigert werden kann (Schunk, 1995a). Zudem ist anzunehmen, dass der Einsatz eines Transparenzpapiers auch das bindendifferenzierte Lernen unterstützt. Denn durch die konkreten Angaben zu Anforderungen und Vorbereitungsmöglichkeiten ist es dem Schüler leichter möglich, sich entsprechend seinen individuellen Fähigkeiten und seinem eigenen Lerntempo auf die Prüfung vorzubereiten.

Der Workshop zum Thema Transparenz begann mit einer Reflexionsübung, in der die Lehrkräfte aufgefordert waren, sich an eine eigene Prüfungssituation zu erinnern und zu überlegen, was sie zur Vorbereitung getan haben. Zudem sollten sie sich gedanklich in eine Prüfungssituation versetzen, in der sie nicht wissen, was sie erwartet. Die Lehrkräfte schilderten dann, was sie einer Person gegenüber empfinden würden, die ihnen die Informationen über Anforderungen und Kriterien der Leistungsbewertung vorenthält. Auch die Auswirkungen der Mitteilung von Informationen über Lehrziel und Bewertung einer Prüfung auf das eigene Lernen, die Motivation und die Leistung wurden diskutiert. Im Anschluss an diese Reflexion wurden die mit Transparenz verbundenen Effekte vorgestellt und die Maßnahme Transparenzpapier erläutert. Mithilfe von bereits in Vorgängerprojekten erstellten Transparenzpapieren sollten die Lehrkräfte im Workshop dann Fachteams bilden und ein eigenes Transparenzpapier für eine in den siebten Klassen anstehende Klassenarbeit entwickeln. Die selbst konzipierten Transparenzpapiere wurden in der Workshop-Gruppe vorgestellt und auf Vollständigkeit überprüft. Da bereits aus Vorgängerprojekten bekannt

war, dass ein Teil der Lehrkräfte auf die Maßnahme des Transparenzpapiers mit Widerständen reagiert, wurden im Workshop dann noch oft angebrachte Einwände diskutiert. Die Projektlehrer wurden aufgefordert, die Transparenzpapiere in den siebten Klassen zu erproben und vor jeder Prüfung einzusetzen.

#### **4.3.3 Lern- und Leistungsraum**

Im dritten Modulworkshop wurde die Maßnahme zur *Trennung von Lern- und Leistungsräumen* erarbeitet. Diese Methode beinhaltet die regelmäßige Einführung sogenannter Lernräume in den Unterricht. In diesen mit den Schülern vereinbarten Zeiten, die bestenfalls durch ein Signal (z. B. Schild, Raumwechsel) für Schüler und Lehrer als Lernraum gekennzeichnet sind, dürfen die Lehrkräfte weder verbal noch schriftlich oder gedanklich Noten vergeben. Die Lehrkraft kann in diesen Lernzeiten die eigenen Schüler und deren individuelle Kompetenzen besser kennenlernen, den Lernraum aber nicht für Bewertungen, jedoch für Diagnosen von Kompetenzen, Lernstrategien und motivationalen Ressourcen der Schüler nutzen. Rückmeldungen der Lehrkraft zum Lerngeschehen der Schüler sollten mittels individueller Bezugsnorm (siehe auch Rheinberg, 1995; Rheinberg & Vollmeyer, 2012) und als attributionales Feedback (siehe auch Möller, 2010; Weiner, 1985) erfolgen und soziale Vergleiche vermeiden. Durch die individuellen und attributionalen Rückmeldungen erhalten die Schüler Zuspruch im Sinne der dritten Quelle der Selbstwirksamkeitserwartung, den verbalen Überzeugungen. Zudem erhöht sich die Wahrscheinlichkeit für die Erfahrung eigener Erfolge, denn kleinste Lernerfolge und -fortschritte sowie der Zuwachs an eigenen Kompetenzen werden sichtbar und die Schüler erleben, dass sich die eigenen Lernbemühungen lohnen. So sollte auch durch den Einsatz des Lernraums die schulische Selbstwirksamkeitserwartung gesteigert werden. Der Lernraum eröffnet darüber hinaus die Chance zum binnendifferenzierten Lernen, sodass jeder Schüler in dem ihm eigenen Tempo Aufgaben bearbeiten kann, die einen für die eigenen Fähigkeiten optimalen, sprich minimal herausfordernden Schwierigkeitsgrad aufweisen. Lernräume sind folglich Zeiten im Unterricht, in denen die Schüler lernen und üben können, ohne angesichts von möglichen Fehlern eine schlechte Benotung befürchten zu müssen. So wird eine Fehlerkultur geschaffen, in der die wahrgenommene Bedrohung von Fehlern und somit auch das schulbezogene Angst- und Stresserleben abnehmen. Auf die Bedeutsamkeit eines angemessenen Umgangs mit Fehlern im Unterricht und die Akzeptanz von Fehlern als natürlichem Bestandteil des Lernens sowie als Lernchance haben Oser und seine Mitarbeiter in verschiedenen Studien hingewiesen (einen Überblick geben Oser und Spsychiger 2005).

Durch eine solche Akzeptanz von Fehlern wird auch die Lernzielorientierung gefördert (Jagacinski, 1992). Voraussetzung für diese Akzeptanz von Fehlern ist nach Chott (1999; 2004) die klare Trennung von Leistungsbeurteilung und Lern- beziehungsweise Übungszeiten im Unterricht, wie sie die Maßnahme zur Trennung von Lern- und Leistungsräumen vorsieht.

Im Workshop wurde die Maßnahme zur Trennung von Lern- und Leistungsräumen sowie die beschriebenen positiven Auswirkungen auf Motivation und Lernen vorgestellt. Danach wurden die Lehrkräfte in fünf Gruppen eingeteilt und aufgefordert, in den Gruppen fünf Fragen zur Trennung von Lern- und Leistungsräumen im Unterricht zu bearbeiten:

- Welche diagnostischen Informationen kann man im Lernraum erhalten, die man im Leistungsraum nicht bekommt?
- Wie kann man sicherstellen, dass die Trennung von Lern- und Leistungsräumen von den Schülern auch wahrgenommen wird?
- Wie kann man sicherstellen, dass sich Schüler im Lernraum nicht „zurücklehnen“?
- Wenn Schüler Fehler im Lernraum machen, welche Rückmeldungen und Reaktionen durch die Lehrkraft sind günstig beziehungsweise ungünstig für eine Lernzielorientierung?
- Wie kann man vermeiden, dass man im Lernraum doch Noten gibt? Was tun, wenn Schüler solche Bewertungen einfordern?

Die Fragen waren auf Postern festgehalten, die im Raum aushingen. Im Sinne einer Stationsarbeit begannen alle Gruppen zeitgleich mit der Bearbeitung einer Frage und hielten ihre Gedanken auf den Postern fest. Nach etwa sieben bis zehn Minuten wechselten die Gruppen zum nächsten Poster und ergänzten dort die Argumente und Anmerkungen der vorherigen Gruppe. Dieses Vorgehen wurde so lange wiederholt, bis jedes Poster von jeder Gruppe einmal bearbeitet worden war und die Gruppen wieder am „Ausgangsposter“ angekommen waren. Die Ergebnisse der Gruppenarbeit wurden dann im Plenum vorgestellt und diskutiert. Auf diesem Weg wurden Ziele des Lernraums sowie Möglichkeiten der Umsetzung und damit verbundene Schwierigkeiten von den Lehrkräften selbst erarbeitet. Jede Lehrkraft sollte den Lernraum im Anschluss an den Workshop im eigenen Unterricht erproben.

#### 4.3.4 Selbstbestimmung

Neben den Bedürfnissen nach sozialer Eingebundenheit und nach Kompetenzerleben stellt laut Deci und Ryan (1985, 1993) auch die *Selbstbestimmung* ein weiteres dem Menschen innewohnendes zentrales Grundbedürfnis dar, dessen Erfüllung eine wesentliche Voraussetzung für Motivation ist. Selbstbestimmtes schulisches Handeln erhöht das eigene Kontrollerleben, wodurch die intrinsische Lernmotivation, das schulische Engagement und auch das emotionale Wohlbefinden in der Schule steigen und Angst- und Stresserleben abnehmen (Deci & Ryan, 2009). Die erlebte Autonomie hat auch einen maßgeblichen Einfluss auf die eigenen Kompetenzüberzeugungen. Wenn sich Schüler als Verursacher ihrer Handlungen wahrnehmen, attribuieren sie Erfolge auf die eigenen Fähigkeiten und Anstrengungen (Boekaerts, 1998; Satow, 1999). Selbst gesetzte Ziele werden zudem mit einer verstärkten Anstrengung verfolgt und auch häufiger erfolgreich erreicht als fremdbestimmte Ziele (Schunk, 1991). Zusammenfassend steigern Lernumwelten, die Eigeninitiative und Wahlfreiheit fördern, das Vertrauen in die eigenen Fähigkeiten und ermöglichen somit den Anstieg von Selbstwirksamkeitserwartungen (Satow, 1999). Durch die Selbst- und Mitbestimmung bei verschiedenen Aufgaben im Unterricht sind darüber hinaus auch die Bindendifferenzierung sowie die Wahl von individualisierten, den Fähigkeiten entsprechenden Aufgaben möglich (Paradies & Linser, 2009), wodurch sich wiederum die Chance auf eigene Erfolgserfahrungen erhöht.

Der vierte Modulworkshop des FOSIS2-Projekts begann mit einer Reflexionsaufgabe zum Thema Selbstbestimmung in der Schule. Die Lehrer erhielten einen Text mit dem Titel „Dicht gedrängt“ (Eisenträger, 2000), in dem ein Arbeitstag beschrieben wird, der dem Schulalltag mit all seinen Zwängen ähnelt (Großraumbüro mit nur der nötigsten Ausstattung; für die Bearbeitung von Aufgaben existieren zeitliche Vorgaben; durch ein Klingelzeichen wird das Ende der Bearbeitungszeit angekündigt; jede Stunde wechseln die Themen der Aufgaben u.v.m). Auf diesem Weg sollten die Lehrkräfte sich in einen Schulalltag der Schüler versetzen und sich darüber bewusst werden, wie wenig Gestaltungsspielräume den Schülern zur Verfügung stehen. Im Anschluss wurden Wahlmöglichkeiten im Unterricht erarbeitet, die den Schülern ein höheres Maß an Selbstbestimmung verschaffen. Wahlmöglichkeiten können hinsichtlich des Themengebietes, des konkreten Inhaltes einer Stunde, der Methode, der Schwierigkeit, der Sozialform, des Zeitpunktes, der Zeitdauer und des Arbeitsortes eingeräumt werden. Im Workshop erarbeiteten die Lehrkräfte konkrete, umsetzbare Wahlmöglichkeiten für ihren jeweiligen Fachunterricht.

### 4.3.5 Kooperatives Lernen

Nach Slavin (1989) ist *Kooperatives Lernen* „a form of classroom organisation in which students work in small groups to help one another learn academic material“ (S. 129). Kooperatives Lernen bezeichnet somit eine Struktur des Unterrichts, bei der sich Schüler gegenseitig unterstützen und gemeinsam Wissen und fachliche sowie soziale Fertigkeiten erwerben. Neben diesem Erwerb von Fachwissen und sozialen Kompetenzen existieren zahlreiche weitere positive Auswirkungen und Korrelate kooperativer Lernformen (im Überblick bei Neber & Fischer, 2010): In sozialer Hinsicht erhöhen sich die generelle Bereitschaft zum prosozialen Verhalten und die Beziehungsqualität unter den Schülern, so dass auch Außenseiter stärker in den Klassenverband integriert werden. Die Tendenz zu sozialen Vergleichen nimmt ab, wohingegen die Wahrnehmung von Gleichheit ansteigt. Darüber hinaus können auch das emotionale Wohlbefinden und motivationale Variablen, wie die Fähigkeit zur Selbstregulation und die Selbstwirksamkeitserwartung, durch kooperative Lernformen verbessert werden (Neber & Fischer, 2010). Im Sinne der zweiten Quelle der Selbstwirksamkeitserwartung, der stellvertretenden Erfahrung, bieten kooperative Lernformen die Möglichkeit, über Beobachtung und Modelllernen stellvertretende und durch Nachahmung dann auch eigene Erfolge zu erfahren (Jerusalem, 2005b). Der Einsatz der im Modul erarbeiteten kooperativen Lernform, des Gruppenpuzzles, bei der leistungsheterogen zusammengesetzte kooperative Lerngruppen gebildet werden, in denen leistungsstarke Schüler die schwächeren gezielt unterstützen können, befördert zudem auch das binnendifferenzierte Lernen (Kress & Pappas, 2013). Zu beachten ist, dass kooperative Lernformen auch misslingen und unerwünschte Effekte nach sich ziehen können. Der sogenannte *Free-Rider-Effekt* bezeichnet eine Gruppenarbeit, bei der sich schwächere Schüler zurücklehnen und den leistungsstärkeren Schülern die Arbeit überlassen. Im umgekehrten Fall, dem sogenannten *Sucker-Effekt*, kann es unter bestimmten Umständen auch passieren, dass sich die leistungsstarken Schüler von den schwächeren ausgenutzt fühlen und in Folge ihre eigene Anstrengung verringern. Auch *statusabhängige Effekte* sind beschrieben worden. Hiernach verringert sich die Lernbereitschaft von statusniedrigen Gruppenmitgliedern. Der *Ganging-Up-Effekt* meint, dass sich die gesamte Gruppe auf die mit der geringsten Anstrengung verbundenen Lösungswege einigt (Neber & Fischer, 2010). Damit es nicht zu diesen negativen Effekten kommt und kooperatives Lernen im positiven Sinne gelingen kann, müssen bestimmte Voraussetzungen erfüllt sein. In der Literatur werden zahlreiche Bedingungen beschrieben, die das Gelingen kooperativer Lernformen begünstigen (Johnson & Johnson, 1995; Renkl & Mandl, 1995; Slavin, 1995). Hier werden nur die

Merkmale dargestellt, die sich in verschiedenen Arbeiten wiederkehrend als förderlich erwiesen haben und die im Rahmen des Workshops mit den Lehrkräften bearbeitet wurden. Eine *positive Abhängigkeit*, bei der alle auf ein gemeinsames Ziel hin arbeiten, das jeder einzelne nur erreichen kann, wenn alle Gruppenmitglieder ihr Ziel erreichen, fördert das Engagement aller Beteiligten. Die *persönliche, individuelle Verantwortlichkeit* als eine weitere wesentliche Bedingung für gelingendes kooperatives Lernen besagt, dass jeder innerhalb der Gruppe für einen Teil der Arbeit verantwortlich ist und die individuelle Leistung identifiziert und bewertet werden kann. Darüber hinaus wird eine *heterogene Gruppenzusammensetzung* z. B. bezüglich sozialer Fähigkeiten, Leistungen, ethnischer Zugehörigkeit oder sozialem Status empfohlen (vgl. auch Kunter & Trautwein, 2013; Neber & Fischer, 2010).

Im Workshop wurde gemeinsam mit den Lehrkräften eine spezielle kooperative Lernmethode praktisch erprobt: das *Gruppenpuzzle* nach Aronson, Blaney, Stephan, Sikes und Snapp (1978). Die Methode des Gruppenpuzzles, auch Jigsaw-Methode genannt, ist 1971 erstmals in einer Schule in Texas zum Einsatz gekommen (Aronson, 2014). In dieser Schule gab es große ethnische Konflikte zwischen Jugendlichen angloamerikanischer, afroamerikanischer und südamerikanischer Herkunft, die mit einer Atmosphäre von Feindseligkeit und Aufruhr verbunden waren. Aronson wurde vom Schulleiter um Hilfe gebeten, das Miteinander zu verbessern. Er nahm zunächst Unterrichtsbeobachtungen vor und stellte fest, dass die Spannungen durch kompetitive Unterrichtsstrukturen befördert wurden. Diese zeichneten sich dadurch aus, dass jeder Schüler um gute Noten und Anerkennung bei den Lehrern kämpfte und Minoritäten wegen mangelnder Sprachkenntnisse oder sozialer Unterschiede benachteiligt wurden. Aronson und seine Mitarbeiter entwickelten daraufhin die Jigsaw-Methode, durch deren Umsetzung sie deutliche Verbesserungen der Atmosphäre und des Miteinanders erzielten: Verglichen mit Kontrollklassen an derselben Schule fühlten sich die Schüler in den kooperativ arbeitenden Klassen geringer benachteiligt und hatten auch gegenüber Minderheiten weniger Vorurteile. Außerdem zeigten sie mehr Selbstsicherheit und Freude an der Schule, weniger Fehlzeiten und bessere Leistungen (Aronson, 2014).

Beim Gruppenpuzzle werden die Schüler in heterogene (in der Ursprungsschule vor allem nach ethnischen Gesichtspunkten gemischte) Gruppen eingeteilt, in denen die Schüler kooperativ Lerninhalte erarbeiten sollen. Wichtige Voraussetzung zur Umsetzung eines Gruppenpuzzles ist ein Thema, das mittels Material erarbeitet werden kann und das sich in



verschiedene sinnvolle Einzeltexte zu einem übergeordneten Thema unterteilen lässt. Mithilfe des in Subthemen untergliederten Materials sind dann zwei verschiedene Gruppenbildungen möglich: Zunächst werden heterogene *Lerngruppen* (auch Basis-, Austausch- oder Home-Gruppen) gebildet, in denen jedes Mitglied seinen Teil des Aufgabenmaterials erhält. Im Anschluss treffen sich die Vertreter eines Textes/Materials in *Expertengruppen* (auch Fokus-Gruppen). Deren Mitgliedern liegt das gleiche Material vor. Sie werden zu Experten, indem sie das Material erarbeiten, dieses diskutieren, Fragen klären, Arbeitsergebnisse festhalten und sich gemeinsam überlegen, wie sie als Experte ihr Wissen an die Lerngruppe weitergeben können. Denn in der nächsten Phase wechseln die Teilnehmer wieder in ihre Lerngruppen und berichten dort nacheinander über ihre Arbeit und ihre Ergebnisse aus den Expertengruppen. Dabei werden offene Fragen geklärt, sodass am Ende jedes Mitglied der Lerngruppe das vorgetragene Teilgebiet verstanden hat und in den Gesamtzusammenhang einordnen kann. Schließlich erfolgt eine Evaluation des Gelernten, indem z. B. ein Test geschrieben wird, der Fragen zu jedem Teilgebiet enthält. Dabei wird deutlich, wie die jeweiligen Experten das Teilgebiet vermittelt und die Teilnehmer das Wissen verarbeitet haben. Wichtig in der letzten Phase ist auch die Kommunikation über die abgelaufene Gruppenarbeit und ihre Bewertung. Nur so können Defizite eruiert und beseitigt werden.

Im Workshop wurde das Gruppenpuzzle zum Thema „Kooperatives Lernen“ durchgeführt. Somit wurde gleichzeitig eine kooperative Lernmethode erprobt und die theoretischen Hintergründe des kooperativen Lernens sowie die Auswirkungen und die förderlichen Rahmenbedingungen erarbeitet. Zudem beschäftigte sich eine Expertengruppe mit verschiedenen kooperativen Lernmethoden und eine weitere mit der Rolle der Lehrkraft bei dem Einsatz solcher Methoden. Mithilfe eines Multiple-Choice-Tests wurde das erarbeitete Wissen zum kooperativen Lernen im Anschluss an das Gruppenpuzzle überprüft.

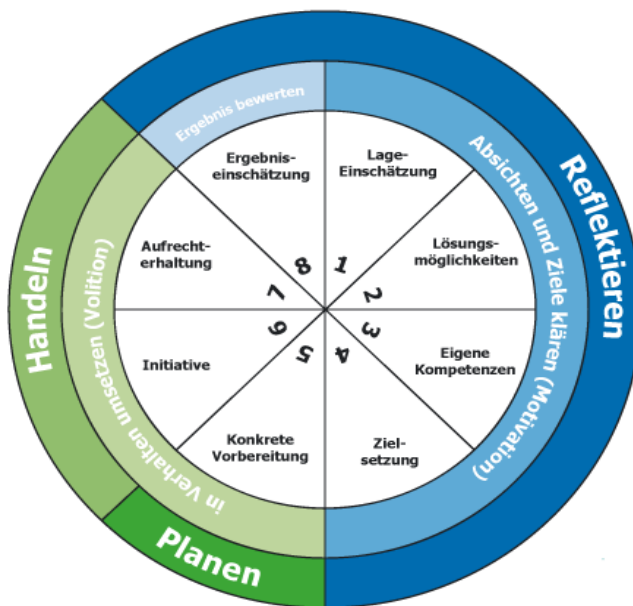
#### **4.3.6 Selbstregulation**

Nach Weinert (1982) sind Lernformen *selbstreguliert*, wenn der Lernende „die wesentlichen Entscheidungen, ob, was, wann, wie und woraufhin er lernt, gravierend und folgenreich beeinflussen kann“ (S. 102). Selbstorganisiertes Lernen und problemlösendes Handeln im alltäglichen Lebensvollzug gelten als Basiskompetenzen, die in der Schule vermittelt werden sollen (Landmann, Perels, Otto & Schmitz, 2009; Traub, 2003). Insbesondere in Lernsituationen, die nicht mit einfachen Routinen bewältigt werden können, benötigen die Schüler Fähigkeiten zur selbstständigen Steuerung des eigenen Lernens (Landmann et

al., 2009). Selbstreguliertes Lernen wirkt sich positiv auf die schulischen Leistungen und auf den Einsatz von Lernstrategien aus (Dignath, Buettner & Langfeldt, 2008). In motivationaler Hinsicht hat sich gezeigt, dass der Einsatz von Selbstregulationsstrategien bei Schülern mit einer positiven Veränderung von Wertvorstellungen, Zielen und Überzeugungen einhergeht. So verbessern sich auch die eigenen Kompetenzüberzeugungen durch den Gebrauch von Selbstregulationsfähigkeiten (Acee & Weinstein, 2010). Zentrale Prozesse der Selbstregulation sind das Setzen, Verfolgen und Bewerten von Lern- und Leistungszielen (Zimmerman, 2000). Gerade diese Prozesse der Zielverfolgung und der Evaluation der Zielerreichung helfen dabei, eigene Erfolge zu realisieren, wodurch die Selbstwirksamkeit gesteigert wird (Schunk, 1995a). Umgekehrt setzen sich hoch selbstwirksame Personen auch eher herausfordernde Ziele und entwickeln Strategien, die ihnen helfen, diese Ziele zu erreichen (Krapp & Ryan, 2002). Sie investieren mehr in selbstregulatorische Prozesse und evaluieren Handlungsprozesse häufiger als Personen mit einer niedrigeren Selbstwirksamkeitserwartung (Usher & Pajares, 2008). Im Gegensatz dazu verringert sich die Fähigkeit zum selbstregulierten Lernen durch Prüfungsangst. Prüfungsängstliche neigen eher dazu, bei der Zielsetzung, Steuerung und Bewertung des eigenen Lernens Anweisungen von Autoritätspersonen Folge zu leisten (Pekrun & Götz, 2006). Insgesamt lässt sich die Fähigkeit zum selbstregulierten Lernen als „Schlüsselkompetenz“ für den schulischen Alltag bezeichnen (Landmann et al., 2009, S. 50), die auch im Rahmen des FOSIS2-Projekts trainiert wird und bei den Interventionsschülern die schulische Selbstwirksamkeitserwartung und durch die Erfahrung von Kontrolle und erfolgreicher Selbstregulation auch Prüfungsangst und Stresserleben positiv beeinflussen soll.

In Anlehnung an das Rubikonmodell (z. B. Gollwitzer, 1990; Heckhausen, Gollwitzer & Weinert, 1987) wurde für das sechste Modul das sogenannte Proaktive Handlungsregulationsmodell entwickelt, das aufzeigt, welche selbstregulatorischen Fertigkeiten beim Handeln erforderlich sind, und praktische Anhaltspunkte für eine Förderung dieser Fertigkeiten bietet (siehe Abbildung 4). Das Modell unterteilt den Handlungsprozess in acht Phasen: Lageeinschätzung des Problems (1), Brainstorming prinzipieller Lösungsmöglichkeiten (2), Einschätzung der eigenen Kompetenzen und Ressourcen in Bezug auf die gesammelten Lösungsmöglichkeiten (3), Entscheidung für einen Lösungsweg und Zielformulierung (4), konkrete Planung (5), vorbereitende Handlungen (6), Durchführung der eigentlichen Handlung (7) und Bewertung des Ergebnisses (8). Durch die Förderung der Internalisierung dieser Phasen soll der Schüler einen vorausschauenden Umgang mit komplexen An-

forderungen sowie einen strukturierten und nicht überfordernden Problemlöseprozess anstreben.



**Abbildung 4: Das Proaktive Handlungsregulationsmodell**

Im Workshop wurde den Lehrkräften zunächst das Konzept des selbstregulierten Lernens und dessen Bedeutsamkeit für den schulischen Kontext vermittelt. Im Anschluss wurde das Proaktive Handlungsregulationsmodell vorgestellt und jede Phase ausführlich besprochen. Die Lehrkräfte sollten dann Probleme und Schwierigkeiten sammeln, mit denen ihre Schüler im schulischen Alltag konfrontiert werden und die sie nicht mit gewohnten Routinen bewältigen können (z. B. ein Referat halten, einen Praktikumsplatz finden, die Organisation eines Klassenausflugs). In einem nächsten Schritt sollten die Lehrer für eine der schwierigen Aufgaben ausarbeiten, was bei der Bearbeitung im Sinne des Proaktiven Handlungsregulationsmodells alles bedacht werden muss. Schließlich wurden die Lehrkräfte gebeten, sich für eine Partnerarbeit zusammenzufinden. Zu zweit sollten sie sich vorstellen, ein Schüler käme mit einer schulischen Anforderung auf sie zu, um Rat zu suchen. Für jede Phase des Handlungsregulationsmodells sollten die Lehrkräfte nun offene Fragen entwickeln, durch die sie den Schüler bei der Selbstregulation des Problems unterstützen könnten. Dabei sollten die Lehrer dem Schüler unter keinen Umständen Lösungen vorgeben, sondern vielmehr Fragen entwickeln, die dem Schüler Hilfe zur Selbsthilfe geben, sodass er durch Reflexion auf eigene Ideen kommen kann. Im Rahmen eines abschließenden Quiz konnten die Lehrkräfte ihr erworbenes theoretisches und praktisches Wissen zum Thema Selbstregulation überprüfen.

#### **4.4 Zusammenfassung der Projektbeschreibung**

Bei dem FOSIS2-Projekt handelt es sich um eine quasi-experimentelle Längsschnittstudie, im Rahmen derer Lehrkräfte von siebten und achten Jahrgängen dreier Berliner Oberschulen als Mediatoren geschult wurden. Durch ein verändertes Unterrichtsverhalten sollen die Lehrkräfte die motivationalen und sozialen Ressourcen und Risikofaktoren ihrer Schüler positiv beeinflussen, wobei der Fokus dieser Arbeit auf den motivationalen Effekten des Projekts liegt. Im Sinne der Unterscheidungen zur Einordnung von motivationalen Förderprogrammen von Rheinberg und Krug (2005b; siehe Kapitel 1) lässt sich das FOSIS2-Programm zusammenfassend durch folgende Merkmale charakterisieren: Die sechs Module beinhalten sowohl Maßnahmen zur Veränderung intrapersonaler Merkmale der Schüler (z. B. Aufbau von Kompetenzüberzeugungen und Fähigkeiten zur Selbstregulation) als auch Maßnahmen zur Veränderung von Kontextbedingungen (z. B. Verbesserung des Klassenklimas, Erhöhung der Transparenz im Unterricht). Das Programm ist multidimensional angelegt und soll gleichzeitig Ressourcen und Risikofaktoren für die Motivation verändern.

Nachfolgend werden die konkreten Fragestellungen und die zugehörigen Hypothesen der vorliegenden Untersuchung ausgeführt.

## 5 Fragestellungen

Ziel dieser Arbeit ist die Evaluation der Wirksamkeit des FOSIS2-Programms: Zunächst wird die kumulative Gesamtwirkung aller FOSIS2-Module über beide Projektjahre hinweg untersucht und im Anschluss geprüft, ob sich jahresspezifische Effekte zeigen. Schließlich soll beispielhaft für das erste Projektjahr analysiert werden, ob die Programmwirksamkeit mit der Programmumsetzung zusammenhängt. Nachfolgend werden zunächst das antizipierte Wirkmodell des FOSIS2-Projekts beschrieben und anschließend die Fragestellungen sowie die Forschungshypothesen erläutert.

### 5.1 Wirksamkeitsevaluation

Abbildung 5 veranschaulicht die angenommenen Wirkungen des FOSIS2-Programms. Auf der linken Seite der Abbildung 5 sind die Module des FOSIS2-Projekts in chronologischer Reihenfolge aufgeführt. Förderstrategien aus bereits eingeführten Modulen sollten die Lehrkräfte beibehalten und deren Umsetzung neben den neu hinzukommenden Maßnahmen fortsetzen. Im Laufe der beiden Projektjahre wurde das Set an erlernten und im Unterricht möglichst dauerhaft praktizierten Maßnahmen also sukzessive erweitert. Eine modulspezifische Evaluation, bei der die Wirkungen der einzelnen Module untersucht werden, ist deshalb nur schwer möglich. Stattdessen muss von einer kumulativen Wirkung der FOSIS2-Maßnahmen ausgegangen werden. Ziel dieser Arbeit ist folglich die Evaluation der globalen Programmwirksamkeit über die Module des ersten und des zweiten Projektjahres hinweg. Die Zielkriterien, auf die sich die Förderstrategien kumulativ auswirken sollen, sind in Abbildung 5, rechts neben den Modulen dargestellt. Die Kriterien wurden von den Modellen und Konzepten abgeleitet, die den theoretischen Hintergrund der Fördermaßnahmen des FOSIS2-Programms bilden (siehe auch Kapitel 2). Erwartet wird, dass durch den Einsatz der FOSIS2-Förderstrategien zunächst *proximale*, das Unterrichtsgeschehen betreffende Kriterien beeinflusst werden. Bei diesen proximalen Variablen handelt es sich um situative oder verhaltensnahe Konstrukte, die Ressourcen und Risiken der Lernumwelt beschreiben. Wenn die FOSIS2-Förderstrategien langfristig und kontinuierlich umgesetzt werden, ist zu erwarten, dass sich auch günstige Veränderungen in *distalen*, stabilen, also eher dispositionalen, personalen Ressourcen und Risikofaktoren zeigen. Die verschiedenen Zielkriterien, für die eine kumulative Wirkung der FOSIS2-Fördermaßnahmen angenommen wird, lassen sich inhaltlich entweder dem Bereich Motivation oder dem Bereich Sozialklima zuordnen. In der vorliegenden Untersuchung liegt der Fokus auf den motivationa-

len Effekten des Programms. Deshalb wurden die Zielkriterien aus dem Bereich Motivation in Abbildung 5 blau unterlegt und die für diese Variablen angenommenen Hypothesen werden im folgenden Abschnitt näher erläutert.

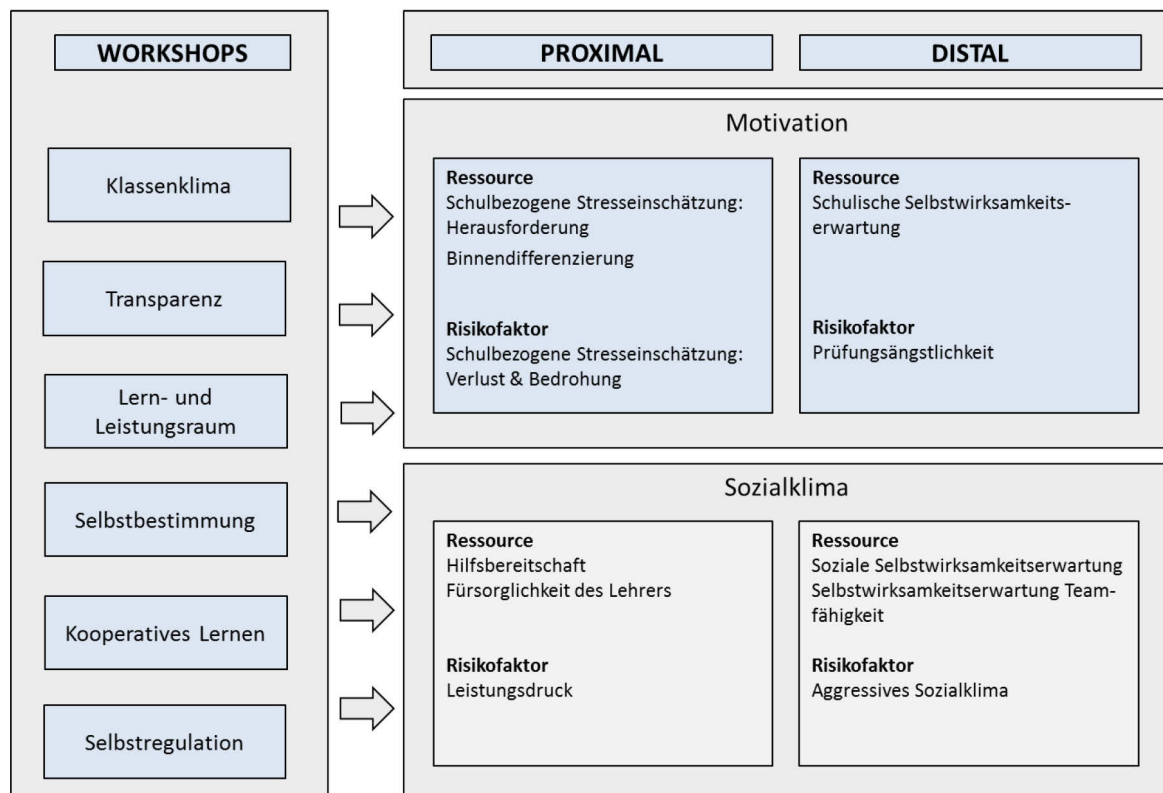


Abbildung 5: Wirkmodell des FOSIS2-Programms

### 5.1.1 Trägt die Intervention zu einer günstigeren Entwicklung der motivationalen Ressourcen und Risikofaktoren bei?

Im Hinblick auf die Wirksamkeit des FOSIS2-Programms wird in der Arbeit folgende Fragestellung untersucht: Trägt die Intervention zu einer günstigeren Entwicklung der motivationalen Ressourcen und Risikofaktoren bei? Eine zentrale personale Ressource für die Motivation ist die Selbstwirksamkeitserwartung nach Bandura (1997), die in Abschnitt 2.2 ausführlich erläutert wurde. Speziell im Kontext Schule ist die für diesen Bereich spezifische schulische Selbstwirksamkeitserwartung ein besonders protektiver motivationaler Faktor. Denn im Gegensatz zu Schülern mit niedriger schulischer Selbstwirksamkeitserwartung zeigen Schüler mit höherer schulischer Selbstwirksamkeitserwartung bei gleicher Fähigkeit eine größere Anstrengung und Ausdauer, ein höheres Anspruchsniveau und ein effektiveres Zeitmanagement sowie größere strategische Flexibilität beim Problemlösen, bessere Leistungen, eine realistischere Einschätzung der Güte ihrer eigenen Leistungen und selbstwertförderlichere Ursachenzuschreibungen (Bandura, 1997; Jerusalem, 2005b;

Usher & Pajares, 2008). Wichtigste Quelle für den Aufbau von Selbstwirksamkeitserwartungen ist nach Bandura (1997) die Erfahrung eigener Erfolge. Aber auch durch stellvertretende Erfahrungen, durch verbale Ermutigung oder das Erleben positiver emotionaler Gefühle kann die Selbstwirksamkeitserwartung aufgebaut werden. Der Zusammenhang zwischen den Maßnahmen des FOSIS2-Programms und den Quellen der Selbstwirksamkeitserwartung wurde in Abschnitt 4.3 dargestellt. Das FOSIS2-Programm verfolgt das globale Ziel, die schulische Selbstwirksamkeit der Schüler zu steigern beziehungsweise den häufig zu beobachtenden Abfall dieser zentralen personalen Ressource nach dem Übergang auf die weiterführende Schule (Wigfield et al., 2011) abzumildern. Da die Selbstwirksamkeitserwartung eine vergleichsweise stabile, eher dispositionale Ressource darstellt (Wigfield et al., 2011), ist ein Anstieg der schulischen Selbstwirksamkeitserwartung jedoch erst nach einer langfristigen, kontinuierlichen Umsetzung der FOSIS2-Fördermaßnahmen zu erwarten. Deshalb ist die schulische Selbstwirksamkeitserwartung in Abbildung 5 als distales Zielkriterium des Programms markiert. Wesentlicher Risikofaktor für die Motivation ist die Prüfungsängstlichkeit, die im Abschnitt 2.3 ausführlich beschrieben wurde. Im Hinblick auf Motivation, Lernen und Leistung zeigen sich durchweg negative Auswirkungen von Prüfungsängstlichkeit (Pekrun & Götz, 2006). Angst und Ängstlichkeit in der Schule können auch die Selbstwirksamkeitserwartung verringern (Schunk, 2003; Usher & Pajares, 2008). Der Zusammenhang zwischen den Förderstrategien der einzelnen Module und der Reduktion von Prüfungsängstlichkeit wurde in Abschnitt 4.3 referiert. Die FOSIS2-Fördermaßnahmen verfolgen das Ziel, einem Anstieg der Prüfungsängstlichkeit entgegenzuwirken beziehungsweise diese zu verringern. Da auch die Prüfungsängstlichkeit einen eher stabilen personalen Risikofaktor darstellt, wird die Reduktion ebenfalls erst nach einer langfristigen und stetigen Umsetzung der Maßnahmen erwartet und das Konstrukt ist deshalb im Wirkmodell (Abbildung 5) unter den distalen Zielkriterien aufgeführt.

Ein proximales Zielkriterium, für das eine Wirkung der FOSIS2-Fördermaßnahmen erwartet wird, ist das schulbezogene Stresserleben. Anhaltender Schulstress kann sich negativ auf die Motivation und das Lern- und Leistungsverhalten auswirken (Hampel et al., 2008). Im Sinne der primären Stressbewertung nach Lazarus und Folkman (1984) werden hier die wahrgenommene schulbezogene Herausforderung und das Bedrohungs- und Verlusterleben untersucht. Erstere stellt eine Ressource für die Motivation dar und soll durch den Einsatz der FOSIS2-Förderstrategien stabilisiert beziehungsweise gesteigert werden. Die beiden Letzteren hingegen sind motivationale Risikofaktoren, die sich im Laufe der beiden Projektjahre bei den Interventionsschülern günstiger entwickeln sollen als bei den Kon-

trollschülern. Ein weiteres proximales Zielkriterium, das durch den Einsatz der FOSIS2-Förderstrategien günstig beeinflusst werden soll, ist die von den Schülern wahrgenommene Binnendifferenzierung im Unterricht als eine externe Ressource der Lernumwelt (Fraser, 1982). Die FOSIS2-Förderstrategien sollen nämlich nicht nur zu einer Veränderung von Bewertungen, Einstellungen und Überzeugungen der Schüler beitragen, sondern auch bei den Lehrkräften zu einer Veränderung von Unterrichtsstrategien führen, die kein expliziter Bestandteil der FOSIS2-Fördermaßnahmen sind.

Zusammenfassend wird bei den Interventionsschülern verglichen mit den Kontrollschülern über die beiden Projektjahre hinweg eine günstigere Entwicklung in den motivational relevanten Zielkriterien des FOSIS2-Programms erwartet. Da für sämtliche motivationale Variablen nach dem Übergang von der Grund- zur weiterführenden Schule über alle Schultypen hinweg eine Verschlechterung zu verzeichnen ist (Wigfield et al., 2011), werden in der Arbeit nicht nur ein Anstieg der Ressourcen und eine Reduktion der Risikofaktoren als günstige Entwicklungen bewertet, sondern auch bereits die Abmilderung negativer Verläufe. Wird beispielsweise bei den Interventionsschülern im Vergleich zur Kontrollgruppe eine signifikant geringere Abnahme der Ressourcen oder eine weniger ausgeprägte Zunahme der Risikofaktoren ermittelt, so kann dies ebenfalls als positive Entwicklung interpretiert werden.

### **5.1.2 Gibt es jahresspezifische Effekte?**

Die geschilderten Hypothesen gelten gleichermaßen für das erste und das zweite Projektjahr. Denn die Lehrkräfte wurden angehalten, die Maßnahmen einmal eingeführter Module neben neu hinzukommenden Modulen weiterhin in ihren Unterricht zu implementieren und die Förderstrategien somit langfristig und regelmäßig umzusetzen. In diesem Sinn ist im Verlauf beider Projektjahre von einer kumulativen Gesamtwirkung aller Module auszugehen. Aufgrund der unterschiedlichen Modulinhalt und eventuell entstehender Wechselwirkungen der verschiedenen Maßnahmen lässt sich im Vorfeld jedoch nicht vollkommen ausschließen, dass sich entgegen der formulierten Hypothesen keine kontinuierliche Entwicklung in den Zielkriterien zeigt. Es könnten beispielsweise jahresspezifische Effekte existieren und die Interventionsschüler in einem Jahr eine Zunahme in einem Kriterium erleben und im anderen Jahr eine Stagnation in derselben Variable. Deshalb soll nicht nur die lineare Veränderung über die beiden Projektjahre hinweg, sondern auch die diskontinuierliche, jahresspezifische Veränderung untersucht werden. Hierbei handelt es sich um ein rein exploratives (nicht hypothesengeleitetes) Vorgehen.



## 5.2 Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation

Durch eine mangelhafte Programmumsetzung kann die Wirkung einer Intervention geschwächt werden (Beelmann & Karing, 2014; Durlak, 2013; Durlak & DuPre, 2008; siehe auch Abschnitt 3.2). Gerade im schulischen Kontext erweist sich die Implementation, z. B. durch Zeitprobleme oder eine fehlende Unterstützung seitens der Schulleitung, als besonders schwierig (Payne, 2009; Payne et al., 2006). In FOSIS2-Vorgängerprojekten konnten die genannten Schwierigkeiten bestätigt werden. Die Untersuchungen legen eine Adaption der Intervention im Hinblick auf eine verbesserte Umsetzbarkeit nahe und weisen auf die Bedeutsamkeit der Evaluation der Implementation hin (Drössler, Jerusalem & Mittag, 2007; Drössler, Röder & Jerusalem, 2007; Röder et al., 2010). Diesen Empfehlungen wurde im FOSIS2-Projekt Rechnung getragen, indem die FOSIS2-Förderstrategien modifiziert und vielfältige Informationen zum Implementationsgeschehen erhoben wurden. In der Arbeit sollen die Einflüsse eines unterschiedlichen Implementationsausmaßes auf die Programmwirkungen analysiert werden.

### 5.2.1 Hängt die Wirksamkeit der Intervention mit dem von den Schülern wahrgenommenen Ausmaß der Implementation zusammen?

Beispielhaft für das erste Projektjahr<sup>25</sup> wird konkret folgende Fragestellung untersucht: Hängt die Wirksamkeit der Intervention mit dem von den Schülern wahrgenommenen Ausmaß der Implementation zusammen? Die Daten für die wahrgenommene Implementation liegen nur für die Interventionsschüler vor. Anders als bei den vorangegangenen Fragestellungen handelt es sich bei den Analysen zum Einfluss der Implementation folglich nicht um Interventions- und Kontrollgruppenvergleiche. Die Ergebnisse können somit nur als erste Hinweise auf mögliche Zusammenhänge zwischen dem Ausmaß der Implementation und der Programmwirksamkeit bewertet werden. Da das Ausmaß der wahrgenommenen Implementation modulspezifisch erfragt wird, können die Ergebnisse außerdem erste Anzeichen für modulspezifische Wirkungen liefern. Auch diese sollten aufgrund der fehlenden Kontrollgruppe und der möglichen Konfundierung mit zuvor eingeführten Modulen höchstens zur Hypothesenbildung für künftige weiterführende Projekte herangezogen und nicht als modulspezifische Effekte interpretiert werden.

---

<sup>25</sup> Eine Begründung der Auswahl des ersten Projektjahres folgt in Abschnitt 6.5.4.

## 6 Methode

In diesem Kapitel wird zunächst die Durchführung des FOSIS2-Projekts dargestellt. Im Anschluss werden die verschiedenen, den jeweiligen Fragestellungen zugrundeliegenden Untersuchungs-Designs, die Stichprobe und auch die in der vorliegenden Studie verwendeten Erhebungsinstrumente beschrieben. Das Kapitel schließt mit einer ausführlichen Schilderung des methodischen Vorgehens, der Mehrebenenanalyse. Dedrick et al. (2009) bemängeln in ihrem Review zur Anwendung von Mehrebenenanalysen, dass in den meisten Studien, die diese Methode gebrauchen, nicht angemessen erklärt wird, welche Voraussetzungen der Methode erfüllt waren, welches Vorgehen gewählt, welche Modelle mit welchen Zufallseffekten genau geschätzt und welche Indizes zur Bewertung der Modellgüte herangezogen wurden. Deshalb erfolgt hier zuerst eine Begründung der Wahl der Methode (Mehrebenenanalysen) und anschließend werden das in der vorliegenden Untersuchung gewählte Vorgehen und die zur Analyse der Fragestellung entwickelten Modelle genau erläutert. Eine Darstellung der Voraussetzungen für den Einsatz der Methode sowie des allgemein üblichen Vorgehens bei der Mehrebenenmodellierung findet sich im Anhang C.

### 6.1 Durchführung des FOSIS2-Projekts

Der zeitliche Verlauf des FOSIS2-Projekts und auch die Modulinhalte des FOSIS2-Programms wurden bereits in Kapitel 4 beschrieben. Nachfolgend werden die Auswahl der Zielgruppe und die Rekrutierung der Stichprobe sowie die Projekt-Durchführung geschildert.

#### 6.1.1 Auswahl und Rekrutierung der Stichprobe

Bei der Wahl von Siebt- und Achtklässlern als Zielgruppe für das FOSIS2-Programm handelt es sich um eine bewusst getroffene Entscheidung, die zum einen auf Vorerfahrungen mit Interventionen für ähnliche Altersgruppen aus Vorgängerprojekten basiert und zum anderen entwicklungspsychologisch begründet ist. Nach dem Übergang auf die weiterführende Schule wird bei Schülern in allen motivational relevanten Ressourcen ein allgemeiner Abfall beobachtet (Wigfield et al., 2011). In Berlin findet der Wechsel auf die weiterführende Schule in der Regel erst nach sechs Jahren Grundschule statt. Um der Reduktion der schulischen Motivation vorzubeugen und gleichzeitig mit Beginn einer neuen Schule soziale Ressourcen und Kompetenzen zu stärken, wurde das FOSIS2-Projekt gezielt in den siebten Jahrgängen – in Berlin also direkt nach dem Wechsel auf die weiterführende Schu-

le – durchgeführt. Damit die Schüler der siebten Jahrgänge die durch die FOSIS2-Interventionen erzielten Veränderungen in sämtlichen Fächern erfahren konnten, sollten möglichst alle Lehrkräfte an den Workshops teilnehmen, die in den siebten Jahrgängen der Interventionsschulen unterrichteten.

Im Abschnitt 4.1 wurde bereits darauf hingewiesen, dass in den Vorgängerprojekten eine teilweise unzureichende und mangelhafte Implementation der Unterrichtsstrategien aufgefallen war. Die teilnehmenden Schulen zeigten eine zu geringe Bereitschaft zu einer verbindlichen Projektteilnahme und ein Teil der Lehrkräfte eine geringe Durchhaltungsmotivation. Bei der Auswahl der FOSIS2-Interventionsschulen wurden deshalb bestimmte schulbezogene Merkmale beachtet, die den beschriebenen Problemen entgegenwirken sollten:

- Die Interventionsschulen sollten ein eigenes Interesse an der Projektteilnahme bekunden und unter Wahrung ihrer eigenen Schulkultur hinter den Programmzielen stehen können.
- Es sollte sich um besonders engagierte, für Innovationen offene Schulen handeln, deren Offenheit durch bereits bestehende Projekte oder auch die Teilnahme an Weiterbildungen sichtbar war.
- Der Lehrplan sollte über die zwei Projektjahre für die Interventionsklassen die gleichen unterrichtenden Lehrkräfte vorsehen, sodass seitens der Lehrkräfte eine kontinuierliche Projektteilnahme über zwei Jahre gewährleistet werden konnte.

Auf der Basis der Kenntnis von Berliner Schulen durch Vorgängerprojekte, mittels des Wissens von Schulexperten aus der Berliner Senatsverwaltung und aufgrund von Empfehlungen der Unfallkasse Berlin wurden geeignete Interventionsschulen identifiziert und die jeweiligen Schulleitungen kontaktiert. In ersten Rekrutierungsgesprächen wurden die in Abschnitt 4.1 dargestellten Teilnahmebedingungen mit den Schulleitungen besprochen und in Kooperationsverträgen schriftlich festgehalten, die von allen Projektpartnern (Interventionsschule, Humboldt-Universität zu Berlin, Unfallkasse Berlin) unterzeichnet wurden. Insgesamt konnten drei Berliner Schulen – eine Realschule und zwei Gesamtschulen mit gymnasialer Oberstufe – für die Teilnahme am FOSIS2-Projekt gewonnen werden.

Zur Akquise der Kontrollschulen wurden die Schulleitungen der Interventionsschulen gebeten, innerhalb ihres eigenen Berliner Bezirks jeweils eine Schule als Kontrollschule zu benennen, die der jeweiligen Interventionsschule in Größe, Zusammensetzung der Schüler- und Lehrerschaft, in Schulkultur und anderen strukturellen Merkmalen ähnelt. In den Kontrollschulen wurden die FOSIS2-Förderstrategien nicht eingeführt. Die Schüler und Lehrer

der siebten und achten Jahrgänge beteiligten sich nur an den drei großen standardisierten Befragungen vor Beginn des Projekts, nach Ende des ersten Projektjahres und nach beiden Projektjahren (in Abbildung 2 = T1 bis T3). Um die Bereitschaft zur Teilnahme an den Befragungen zu erhöhen, wurde den Kontrollschulen ein monetärer Anreiz geboten. Sollten 90 % der in den siebten und im zweiten Projektjahr achten Klassen unterrichtenden Lehrkräfte einer Schule an allen drei Befragungen teilnehmen, würde die Schule 1000 Euro, bei einer Teilnahme von 80 % noch 500 Euro erhalten. Zudem erhielt jeder Lehrer der Kontrollschulen pro Befragungsteilnahme 10 Euro.

### **6.1.2 Durchführung der Workshops, Beratungen und Befragungen**

Das FOSIS2-Projekt umfasst die sechs beschriebenen Modulworkshops, die verschiedenen Beratungselemente, mittels derer die Programmumsetzung in den Fachunterricht erleichtert werden sollte, und unterschiedliche Befragungen der Schüler und Lehrer aus den Interventions- und Kontrollschulen. Nachfolgend wird die Durchführung der verschiedenen Bestandteile des FOSIS2-Projekts dargestellt.

#### ***Durchführung der Workshops***

Die sechs Workshops des FOSIS2-Programms wurden so über die beiden Projekt-Schuljahre verteilt, dass den Lehrkräften nach jedem Workshop etwa drei Monate Zeit für die Umsetzung der neu erlernten Förderstrategien und die Fortführung der in vorangegangenen Workshops erarbeiteten Maßnahmen blieben. Durch die Schulferien wurden diese Implementationsphasen teilweise unterbrochen, sodass die Workshops nicht exakt im Abstand von drei Monaten aufeinander folgten. Es wurde jedoch darauf geachtet, dass die Workshops in den einzelnen Schulen zur gleichen Zeit – wenn möglich in derselben Woche – stattfanden. Um die Workshops in allen Interventionsschulen möglichst identisch zu gestalten, wurden diese vom selben Trainer, von Professor Matthias Jerusalem, dem Inhaber des Lehrstuhls für Pädagogische Psychologie und Gesundheitspsychologie der Humboldt-Universität zu Berlin, durchgeführt. Die Workshops dauerten je nach Komplexität des Moduls zwischen eineinhalb und drei Stunden und liefen in der Regel nach dem gleichen Schema ab: Jeder Workshop begann mit einer theoretischen Einführung in das Thema. Vorgestellt wurden jene Theorien und Konzepte, auf deren Basis die Förderstrategien entwickelt wurden. Im Anschluss folgte eine Phase, in der die konkrete Umsetzung der Förderstrategien in Lehrer-Teams gemeinsam geplant und erste Schritte für den eigenen Unterricht erarbeitet und eingeübt wurden. Dabei wurden auch mögliche Schwierigkeiten

bei der Umsetzung besprochen und bereits erste Lösungsmöglichkeiten für den Umgang mit diesen Problemen entwickelt. Die individuell geplanten Schritte wurden in Form eines Maßnahmenplans schriftlich von den Lehrkräften notiert. In jedem Workshop kamen darüber hinaus unterschiedliche Materialien zum Einsatz: So wurden beispielsweise die Theorien und Förderstrategien in Form einer PowerPoint-Präsentation dargestellt. Diese erhielten die Lehrkräfte in gekürzter Form zu Beginn eines jeden Workshops auch als Ausdruck. Zu jedem Modul wurde im Vorfeld des Workshops zudem eine sogenannte Handreichung verfasst, die alle wesentlichen Inhalte und Maßnahmen des Moduls sowie praktische Tipps für die Durchführung und den Umgang mit Schwierigkeiten beinhaltet. Während des Workshops kamen zudem verschiedene Arbeitsblätter zum Einsatz, die als Anleitung für die praktischen Übungen dienten. Alle diese Materialien wurden den teilnehmenden Lehrkräften auch digital auf einer Webseite zur Verfügung gestellt. Hier konnten die Lehrkräfte auch eigene im Rahmen des FOSIS2-Projekts erarbeitete Materialien, wie z. B. Transparenzpapiere, hochladen und austauschen.

### ***Durchführung der Beratungen***

In Vorgängerprojekten hatte sich die Umsetzung der Interventionsinhalte in den Schulalltag für die teilnehmenden Projektlehrkräfte teilweise als schwierig erwiesen (siehe Abschnitt 4.1). Wichtiges Anliegen des FOSIS2-Projekts war es deshalb, ein nachhaltiges Beratungsangebot zu schaffen, durch das die Lehrkräfte bei der Umsetzung der FOSIS2-Maßnahmen in ihren Fachunterricht unterstützt werden konnten.

Dieses Beratungsangebot sah so aus, dass jeder Interventionsschule zunächst eine wissenschaftliche Mitarbeiterin des FOSIS2-Projektteams zugeordnet wurde, die jeweils eine Interventions- und die zugehörige Partnerkontrollschule betreuen sollte. Die Mitarbeiterin war zum einen Ansprechpartnerin für die Koordination von Terminen für Workshops und Befragungen und stand für Rückfragen zur Verfügung. Zum anderen konnten die Interventionsschulen Einzel-, Gruppen-, E-Mail- und Telefon-Beratungen durch die Schulbetreuerin wahrnehmen und es bestand darüber hinaus das Angebot von Unterrichtshospitationen, kurzen Auffrischungen der Workshopinhalte sowie einer unterstützenden Teilnahme der wissenschaftlichen Mitarbeiterin bei Elternabenden und Schulinformationstagen. Die Vielfalt an verschiedenen Beratungselementen wurde geschaffen, um der Komplexität und der Zeitnot im Schulalltag gerecht zu werden. Die Unterstützungs- und Beratungsangebote wurden zeitlich und im Format flexibel den Bedürfnissen der jeweiligen Projektschulen angepasst. Am meisten waren modulspezifische Gruppenberatungen nachgefragt: Ein bis

fünf Mal pro Modul trafen sich Lehrerteams mit ihrer „Schul-Betreuerin“ für ein bis zwei Stunden, um über die Umsetzung der Modulinhalte in ihren Fachunterricht zu sprechen. Die modulspezifischen Gruppenberatungen fanden im Rahmen von klassenübergreifenden teilweise auch fachhomogenen Teamtreffen der Lehrer statt, die ebenfalls durch das FOSIS2-Projektteam angeregt wurden. Durch die Teambildung sollte auch die kollegiale Unterstützung unter den Lehrern gefördert werden.

### ***Durchführung der Befragungen***

Im FOSIS2-Projekt wurden verschiedene standardisierte Befragungen durchgeführt:

- Vor dem ersten Projektjahr, am Ende des ersten und am Ende des zweiten Projektjahres standardisierte Befragungen mit Schülern und Lehrern der Interventions- und der Kontrollschulen (siehe Abbildung 2 = T1 bis T3)
- Standardisierte Wochen-Feedbacks zum Monitoring der Implementation der Förderstrategien mit den Lehrern der Interventionsschulen (siehe Abbildung 2 = F)
- Am Ende der Implementationsphase eines jeden Moduls standardisierte modulspezifische Rückmeldungen und qualitative Interviews zur Implementation der Förderstrategien mit Schülern und Lehrern der Interventionsschulen (siehe Abbildung 2 = M1 bis M3)

Alle Lehrerbefragungen wurden online mit einem Tool von Unipark (QuestBack GmbH, 2014) durchgeführt.<sup>26</sup> Zur Beantwortung der Fragestellungen werden die standardisierten Prä-, Zwischen- und Post-Befragungen mit den Schülern der Interventions- und der Kontrollgruppe sowie die standardisierten modulspezifischen Rückmeldungen der Interventionsschüler genutzt. Die Schüler wurden vor Ort mittels Paper- und Pencil-Erhebungen befragt. Die Durchführung der Umfragen erfolgte innerhalb einer Schulstunde durch Projektmitarbeiter der Humboldt-Universität. Um ein möglichst störungsfreies Ausfüllen der Fragebögen zu ermöglichen, war dabei immer auch eine Lehrperson anwesend. Um den Anreiz für ein sorgfältiges Ausfüllen der Fragebögen und die Teilnahmemotivation zu steigern, wurden während der Befragungen Giveaways der Unfallkasse Berlin verteilt (z. B. Radiergummis, Katzenaugen für Fahrräder, USB-sticks und Traubenzucker). Da sich insbesondere bei den Schülern der Interventionsschulen, die pro Schuljahr an vier bis fünf

---

<sup>26</sup> Die Analyse der Lehrerdaten ist Bestandteil einer anderen im FOSIS2-Projekt angesiedelten Untersuchung. Diese folgt der Annahme, dass durch die FOSIS2-Fortbildungen auch Überzeugungen und Kompetenzen der Lehrkräfte positiv beeinflusst werden, wodurch wiederum auf Seiten der Schüler ein Anstieg von Kompetenzen und Leistungen zu erwarten ist. Denn Merkmale der Lehrperson haben einen bedeutenden Einfluss auf schulische Leistungen (siehe z. B. Blömeke 2009; Hattie 2009).

Befragungen teilnehmen mussten, im Laufe des Projekts eine zunehmende Ermüdung und Reaktanz gegenüber den Befragungen zeigten, wurden die Anreize mit der Zeit gesteigert. So konnten die Schüler neben den Giveaways noch zwei Kinogutscheine pro Klasse gewinnen. Bei der letzten Befragung wurde in den Interventionsklassen schließlich ein H&M-Gutschein im Wert von 25 Euro eingesetzt und in den Kontrollklassen vier Kinogutscheine pro Klasse verlost. Während der ersten Befragung wurde auch deutlich, dass die Schüler für die Beantwortung der Fragen unterschiedliche Bearbeitungszeiten benötigten. Deshalb wurden zusätzlich zu den Anreizen Rätsel an diejenigen Schüler verteilt, die mit dem Ausfüllen des Fragebogens frühzeitig fertig waren. So konnten die Schüler, die zur Beantwortung des Fragebogens mehr Zeit brauchten, ungestört weiter arbeiten.

### **6.1.3 Besondere Begleitumstände**

Beim FOSIS2-Projekt handelt es sich um eine quasi-experimentelle Studie unter schulischen Realbedingungen. Anders als in experimentellen Laborstudien, wo die Zuteilung zur Interventions- und Kontrollgruppe per Randomisierung erfolgt und Störfaktoren kontrolliert werden können, ist es in Feldstudien, wie dem FOSIS2-Projekt, nicht so einfach möglich, den Einfluss von Störfaktoren zu verhindern. Deshalb sollten potentielle Störfaktoren und besondere Begleitumstände während des Projekts beachtet und bei der Ergebnisinterpretation berücksichtigt werden. Solche Begleitumstände und relevante Rahmenbedingungen in den Schulen werden nachfolgend beschrieben.

#### ***Die Berliner Schulreform 2010/2011***

Zum Schuljahr 2010/2011, zu Beginn des zweiten Projektjahres des FOSIS2-Projekts, mussten sich alle Berliner Oberschulen einer Schulreform unterziehen. Aus fünf verschiedenen Typen weiterführender Schulen (Haupt-, Real-, verbundene Haupt- und Real-, Gesamtschule und Gymnasium) wurden zwei weiterführende Schularten: die sogenannte Integrierte Sekundarschule und das Gymnasium. Haupt-, Real- und Gesamtschulen wurden also zu einer Schulform zusammengelegt (Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin, 2014). Im FOSIS2-Projekt waren hiervon vor allem die beiden teilnehmenden Realschulen betroffen. Aber auch die vier Gesamtschulen verzeichneten aufgrund der Schulreform einen stärkeren Zuwachs an Schülern. Da die bereits begonnenen Jahrgänge – also auch die am FOSIS2-Projekt teilnehmenden Klassen – nach dem alten Schulmodell fortgeführt wurden, blieben die Projektklassen von der Schulreform unberührt. Dadurch konnte das FOSIS2-Projekt in organisatorischer Hinsicht ohne Störungen

durch die Schulreform fortgesetzt werden. In den Beratungssitzungen berichteten die Lehrkräfte jedoch immer wieder über mit der Schulreform zusammenhängende Ängste vor zu großen Klassen, verhaltensauffälligen Schülern und einer damit verbundenen Überforderung. Auch wenn sich diese Ängste und Befürchtungen nicht direkt auf die Projektklassen, sondern auf die im Schuljahr 2010/2011 neu gebildeten siebten Klassen bezogen, wäre es möglich, dass die Ängste das FOSIS2-Projekt trotzdem indirekt beeinflusst haben. So könnten die genannten Befürchtungen Auswirkungen auf das Unterrichtsverhalten der Projektlehrer nach sich ziehen und z. B. den Rückgriff auf alte, bekannte Routinen verstärken. Deshalb ist es wichtig, die Schulreform bei der Ergebnisinterpretation, speziell bei der Betrachtung der Implementation der FOSIS2-Maßnahmen, zu berücksichtigen.

### ***Besondere Umstände in den Projektschulen***

Die oben geschilderten Kriterien, anhand derer die Auswahl der Projektschulen erfolgte, wurden weitestgehend eingehalten. Bei allen Interventionsschulen handelte es sich um innovationsfreudige Schulen, die ein eigenes Interesse an der Projektteilnahme bekundeten und sich vertraglich zu einer kontinuierlichen Beteiligung über die zwei Projektjahre verpflichteten. Die Lehrpläne sahen über die zwei Schuljahre hinweg eine stabile Betreuung der Interventionsklassen durch die gleichen Lehrkräfte vor und jede Interventionsschule konnte eine Kontrollschule gewinnen, die der jeweiligen Interventionsschule in Größe, Zusammensetzung der Schüler- und Lehrerschaft, in Schulkultur und anderen strukturellen Merkmalen ähnelte. Alle Schulleitungen zeigten sich in den Rekrutierungsgesprächen äußerst interessiert und kooperativ. Die teilnehmenden Lehrkräfte beklagten jedoch im Laufe des Projekts immer wieder eine zu geringe Unterstützung durch die Schulleitung, insbesondere einen mangelnden Ausgleich, z. B. durch Freistunden, für die Teilnahme am FOSIS2-Projekt und den damit verbundenen zusätzlichen Zeitaufwand. Auch die Bildung von klassenübergreifenden Lehrerteams zur kollegialen Unterstützung bei der Programmumsetzung konnte nicht in allen Schulen optimal umgesetzt werden. In der Schule 1<sup>27</sup> beispielsweise fanden nur unregelmäßig Teamtreffen statt. Dies hing unter anderem mit den strukturellen Rahmenbedingungen der Schule selbst zusammen. Statt eines gemeinschaftlichen Lehrerzimmers existierten für jeden Jahrgang der Schule sogenannte Jahrgangsbüros,

---

<sup>27</sup> Zur Wahrung der Anonymität der Schulen wurden diese durchnummeriert. Mit Schule 1 bis 3 werden die Interventionsschulen bezeichnet, wobei die Schule 3 die Realschule darstellt. Die Kontrollschulen erhalten die Nummern 4 bis 6. Dabei ist die Nummer 4 die Partner-Kontrollschule zur Nummer 1, die Nummer 5 die Partner-Kontrollschule zur Nummer 2 und die Nummer 6 die Partner-Kontrollschule zur Nummer 3. Auch bei der Nummer 6 handelt es sich somit um eine Realschule.



in denen sich die Lehrkräfte aufhalten konnten, die in dem jeweiligen Jahrgang unterrichteten. Da jedoch jeder Lehrer gleichzeitig in unterschiedlichen Jahrgängen unterrichtete, hielten sich die Lehrkräfte je nach Stundenplan in unterschiedlichen Jahrgangsbüros auf. Ein kontinuierlicher Kontakt und Austausch zwischen den Lehrern ist dadurch erschwert. Dies schlug sich dann auch in einem eher unregelmäßigen Treffen von Teams der Projektlehrer der Schule 1 nieder. In den anderen beiden Schulen hingegen gelang es besser, klassenübergreifende Kleinteams zu bilden, die sich im Laufe der zwei Projektjahre in gleicher Zusammensetzung regelmäßig trafen, um über die Umsetzung der FOSIS2-Förderstrategien zu sprechen.

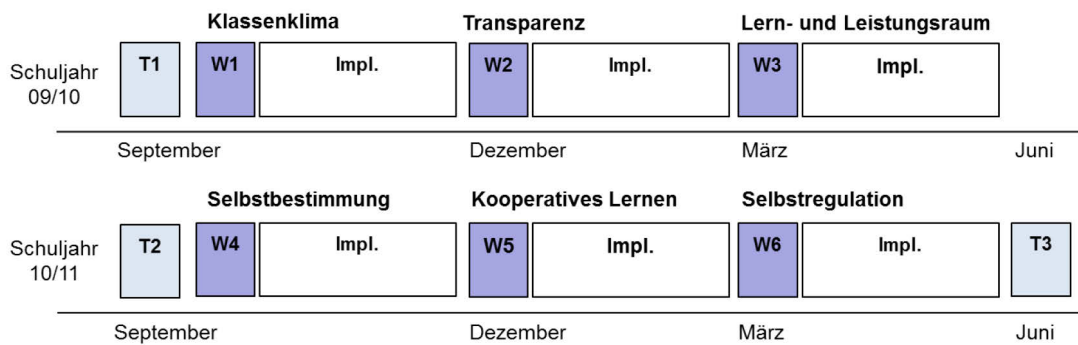
## **6.2 Design**

Bei der Projektbeschreibung (Kapitel 4) ist deutlich geworden, dass im gesamten Verlauf des FOSIS2-Projekts verschiedene Daten zu den Zielkriterien des Programms und auch zur Programmumsetzung erfasst wurden. Der Fokus der vorliegenden Arbeit liegt auf den Schülerdaten. Speziell wird die Wirksamkeit des FOSIS2-Programms im Hinblick auf die motivationalen Effekte auch unter Berücksichtigung der von den Schülern wahrgenommenen Implementation untersucht. Nachfolgend wird das jeweilige Evaluationsdesign, das den Analysen dieser Arbeit zugrunde liegt, näher beschrieben.

### **6.2.1 Design zur Evaluation der Wirksamkeit**

Die Evaluation der Wirksamkeit des FOSIS2-Programms im Hinblick auf die motivationalen Effekte erfolgt durch ein klassisches quasi-experimentelles Kontrollgruppendesign, das in Abbildung 6 veranschaulicht wird. Da die Lehrkräfte angehalten waren, die Förderstrategien aus bereits eingeführten Modulen beizubehalten und deren Umsetzung neben den neu hinzukommenden Maßnahmen fortzusetzen, ist die Evaluation spezifischer Wirkungen einzelner Module erschwert und es können nur die kumulativen Wirkungen der verschiedenen Module, im Sinne einer globalen Programmevaluation, evaluiert werden. Für die Analysen zur Wirksamkeitsevaluation werden die standardisierten Prä-, Zwischen- und Post-Befragungen der Interventions- und der Kontrollschüler herangezogen (in Abbildung 6= T1 bis T3). Mithilfe von Mehrebenenanalysen wird zum einen die lineare Veränderung in den motivational relevanten Zielkriterien untersucht und geprüft, ob sich Interventions- und Kontrollgruppe in dieser linearen Veränderung über beide Projektjahre hinweg bedeutsam unterscheiden. Zum anderen werden sogenannte diskontinuierliche Mehrebenenmodelle modelliert, mithilfe derer in den motivationalen Zielkriterien nach dem ersten (in

Abbildung 6 = T2) und nach dem zweiten Projektjahr (in Abbildung 6 = T3) die Unterschiede zwischen Interventions- und Kontrollgruppe, also die jahresspezifischen motivationalen Effekte, analysiert werden können (eine ausführliche Darstellung des methodischen Vorgehens findet sich in Abschnitt 6.5).

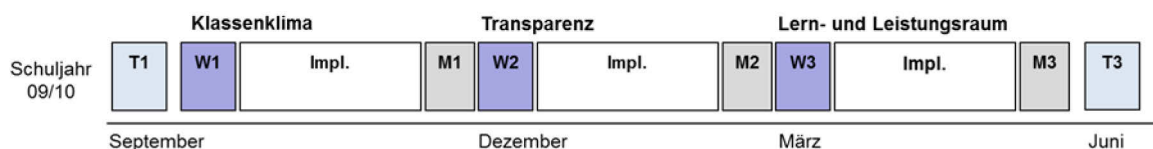


**Abbildung 6: Design zur Evaluation der Wirksamkeit**

*Anmerkungen:* T = Prä-, Zwischen- und Postbefragung, W = Workshop, Impl. = Implementationsphase

## 6.2.2 Design zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation

Die Frage nach der Wirksamkeit des FOSIS2-Programms im Hinblick auf die motivational relevanten Zielkriterien in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler, wird beispielhaft für das erste Projektjahr untersucht. Abbildung 7 veranschaulicht das Design für diese Fragestellung grafisch. Eine Begründung für die Reduktion der Analysen auf das erste Projektjahr findet sich in Abschnitt 6.5.4. Für die Analysen der motivationalen Zielkriterien werden nur der erste und zweite Messzeitpunkt herangezogen (in Abbildung 7 = T1 und T2). Der erste Messzeitpunkt (in Abbildung 7 = T1) wird wiederum als Prä-Befragung, der zweite Messzeitpunkt (in Abbildung 7 = T2) jetzt nicht mehr als Zwischen-, sondern als Post-Befragung behandelt. Als Indikatoren für die Umsetzung der Förderstrategien kommen in dieser Untersuchung die Daten der ersten drei standardisierten modulspezifischen Rückmeldungen zum Einsatz (in Abbildung 7 = M1 bis M3). Diese liegen nur für die Interventionsschüler vor. Deshalb beschränken sich auch die Analysen der T1- und der T2-Befragungen zur Beantwortung der Frage nach der Wirksamkeit in Abhängigkeit von der Implementation auf die Daten der Interventionsschüler.



**Abbildung 7: Design zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation**

Anmerkungen: T = Prä-, Zwischen- und Postbefragung, W = Workshop, Impl. = Implementationsphase

## 6.3 Stichprobe

In diesem Unterkapitel werden zunächst die demografischen Merkmale der Gesamtstichprobe der Schüler und die messzeitpunktspezifische Beteiligung dargestellt. Im Anschluss wird der Umgang mit fehlenden Werten erläutert und schließlich die verschiedenen Analyse-Stichproben beschrieben, die den Berechnungen zugrunde liegen.

### 6.3.1 Beschreibung der Gesamtstichprobe

Die Gesamtstichprobe umfasst Schüler aus 22 siebten und im zweiten Projektjahr achten Klassen dreier Interventions- und drei weiterer Kontrollschulen. Tabelle 1 gibt einen Überblick über die messzeitpunktspezifische Beteiligung und die demografischen Merkmale Geschlecht, Alter und Nationalität.

An der Prä-Befragung nahmen 414 Schüler, an der Zwischen-Befragung 479 und an der Post-Befragung 470 teil, wobei durchschnittlich 20 ( $\pm 2$ ) Schüler pro Klasse befragt wurden.<sup>28</sup> An allen drei Befragungen beteiligten sich etwas mehr Schülerinnen als Schüler (an der ersten Befragung 52.2 % Mädchen, an der zweiten 50.5 % und an der dritten 50.7 %). Das Durchschnittsalter lag zum ersten Messzeitpunkt bei 12 Jahren und vier Monaten, zum zweiten Messzeitpunkt bei 13 Jahren und fünf Monaten und zum dritten Messzeitpunkt bei 14 Jahren und einem Monat. Zu allen drei Messzeitpunkten gaben 71 % bis 75 % der Schüler an, eine deutsche und 15 % bis 19 %, keine deutsche Nationalität zu haben.

<sup>28</sup> Die Angabe der durchschnittlichen Anzahl der Schüler pro Klasse ist wichtig für das in der Studie verwendete methodische Verfahren der Mehrebenenanalyse (siehe auch Abschnitt 6.5.).

**Tabelle 1:** Evaluationsbefragungen – Gesamtstichprobe der Schüler

	T1 (September 2009)	T2 (September 2010)	T3 (Juni 2011)
<b><i>Gesamtstichprobe</i></b>			
Schüler	414	479	470
Klassen (Schulen)	22 (6)	22 (6)	22 (6)
durchschnittliche Anzahl der Schüler pro Klasse	18.8	21.8	21.4
<b><i>Geschlecht</i></b>			
weiblich	216 (52.2%)	242 (50.5%)	238 (50.7%)
männlich	189 (45.6%)	220 (45.9%)	214 (45.5%)
keine Angabe	9 (2.2%)	17 (3.6%)	18 (3.8%)
<b><i>Alter</i></b>			
M (Jahre; Monate)	12;4	13;5	14;1
SD (Monate)	8	10	7
<b><i>Nationalität</i></b>			
Deutsch	309 (74.6%)	359 (74.9%)	337 (71.7%)
andere Nationalität	66 (15.9%)	88 (18.4%)	72 (15.3%)
Keine Angabe	39 (9.5%)	32 (6.7%)	61 (13.0%)

*Anmerkungen:* T = Messzeitpunkt; M = Mittelwert; SD = Standardabweichung

Die Interventions- und die Kontrollgruppe umfassen jeweils 11 Klassen. Tabelle 2 gibt, getrennt für die Interventions- und die Kontrollgruppe, einen Überblick über die Beteiligung an den drei Evaluationsbefragungen. Am ersten Messzeitpunkt nahmen aus den Interventionsklassen 218 Schüler und aus den Kontrollklassen 196 Schüler teil. An der zweiten Befragung beteiligten sich 223 Schüler der Interventions- und 256 der Kontrollgruppe. Die Gesamtheit der 470 Schüler, die den Fragebogen der dritten und letzten Evaluationsbefragung ausgefüllt haben, setzt sich aus 213 Schülern der Interventions- und 257 Schülern der Kontrollklassen zusammen. Von insgesamt 570 Schülern, die an einer der drei Befragungen teilgenommen haben, beteiligten sich 314 an allen drei Befragungen. Von diesen Schülern mit vollständigen Datensätzen sind 157 Schüler aus Interventions- und ebenfalls 157 Schüler aus Kontrollklassen.<sup>29</sup> Die Analysen zur Beantwortung der Frage nach der Wirksamkeit des FOSIS2-Programms werden auf der Basis aller verfügbaren Daten der 570 Schüler durchgeführt (Analyse-SP 1, Tabelle 2). Eine Begründung dieser Stichprobenwahl für die statistischen Analysen erfolgt im Abschnitt 6.3.2.

<sup>29</sup> Die gleichmäßige Verteilung der Schüler mit vollständigen Datensätzen zu allen drei Befragungen in der Interventions- und Kontrollgruppe (jeweils N = 157) ist zufällig entstanden.

**Tabelle 2:** Evaluationsbefragungen – Schülerstichprobe, getrennt nach Interventions- und Kontrollgruppe

	T1 (Sept. 2009)	T2 (Sept. 2010)	T3 (Juni 2011)	T1+T2+T3	T1/T2/T3 (Analyse-SP 1)
<b><i>Interventionsgruppe</i></b>					
Schüler	218	223	213	157	271
Klassen (Schulen)	11 (3)	11 (3)	11 (3)	11 (3)	11 (3)
durchschnittliche Anzahl der Schüler pro Klasse	19.8	20.3	19.4	14.3	24.6
<b><i>Kontrollgruppe</i></b>					
Schüler	196	256	257	157	299
Klassen (Schulen)	11 (3)	11 (3)	11 (3)	11 (3)	11 (3)
durchschnittliche Anzahl der Schüler pro Klasse	17.8	23.3	23.4	14.3	27.2
<b><i>Gesamt (Anzahl Schüler)</i></b>	<b>414</b>	<b>479</b>	<b>470</b>	<b>314</b>	<b>570</b>

*Anmerkungen:* T = Messzeitpunkt; SP 1 = Stichprobe, die den Analysen zur Evaluation der Wirksamkeit zugrunde liegt

Neben den Evaluationsbefragungen werden zur Beantwortung der Frage nach der Wirksamkeit der FOSIS2-Maßnahmen unter Berücksichtigung der von den Schülern wahrgenommenen Implementation auch die modulspezifischen Rückmeldungen (MSR) der Schüler aus dem ersten Projektjahr herangezogen.<sup>30</sup> Da diese als Implementations-Check dienten und die Kontrollschüler nicht zur Implementation befragt werden können, wurden die MSR nur bei den Interventionsschülern eingesetzt. Um die Befragungszeit zu verringern und eine Reaktanz gegenüber den alle drei Monate stattfindenden Rückmeldungen zu vermeiden, wurden in den MSR keine demografischen Variablen erhoben. Tabelle 3 gibt einen Überblick über die Beteiligung der Interventionsschüler an den ersten drei MSR. Aus den insgesamt 11 Interventionsklassen nahmen an der ersten MSR 207 und an der zweiten MSR 212 Schüler teil. Bei der dritten MSR, an der sich 209 Schüler beteiligten, konnte eine Klasse aus organisatorischen Gründen nicht befragt werden.<sup>31</sup> Die Frage nach der Wirksamkeit des FOSIS2-Programms in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler wird beispielhaft für das erste Projektjahr untersucht. Für die Analysen werden die ersten beiden Evaluationsbefragungen und die drei MSR des ersten Jahres herangezogen. Von insgesamt 273 Interventionsschülern, die an mindestens einer der MSR-Befragungen teilgenommen haben, beteiligten sich 130 Schüler an allen drei Be-

<sup>30</sup> Eine Begründung der Auswahl des ersten Projektjahres folgt in Abschnitt 6.5.4.

<sup>31</sup> Die Klassenlehrerin hatte nicht bedacht, dass ihre Klasse an dem vereinbarten Befragungstermin auf einer eintägigen Exkursion war. Da die MSR3 für diese Klasse am vorletzten Schultag vor den Sommerferien geplant war (5. Juli 2010), konnte die Befragung vor den Sommerferien nicht wiederholt werden. Zu Beginn des neuen Schuljahres wurde die zweite große Evaluationsbefragung realisiert. Dadurch konnte die MSR3-Befragung in der einen Klasse auch nach den Ferien nicht erhoben werden, denn eine zeitgleiche Umsetzung beider Befragungen lehnte die Klassenlehrerin ab, weil sonst zu viele Schulstunden aufgrund der Befragungen ausgefallen wären.

fragungen. Der Einfluss der wahrgenommenen Implementation auf die Zielkriterien wird auf der Basis der 130 Schüler mit vollständigen Datensätzen analysiert (Analyse-SP 2, Tabelle 3). Die Entscheidung für diese Stichprobe als Analysestichprobe wird nachfolgend begründet.

**Tabelle 3:** Modulspezifische Rückmeldungen – Schülerstichprobe (nur Interventionsschüler)

	MSR1 (Dez. 2009)	MSR2 (März 2010)	MSR3 (Juni 2010)	T1+T2+ MSR1+2+3 (Analyse-SP 2)	T1/T2/ MSR1/2/3/T2
<b><i>Interventionsgruppe</i></b>					
Schüler	207	212	209	130	273
Klassen (Schulen)	11 (3)	11 (3)	10 (3)	10 (3)	11(3)
durchschnittliche Anzahl der Schüler pro Klasse	18.8	19.2	20.9	13.0	24.8

*Anmerkungen:* T = Messzeitpunkt; MSR = modulspezifische Rückmeldungen; SP 2 = Stichprobe, die den Analysen zur Evaluation der Wirksamkeit in Abhängigkeit von der Implementation zugrunde liegt

### 6.3.2 Umgang mit fehlenden Werten und Wave-Nonresponse-Analysen

In der psychologischen Feldforschung sind die Datensätze oft unvollständig (Lüdtke, Robitzsch, Trautwein & Köller, 2007). Insbesondere bei Längsschnittstudien nehmen Studienteilnehmer an manchen Messzeitpunkten teil und fehlen an anderen (*„wave- oder unit-nonresponse“*). Zusätzlich kann es vorkommen, dass Schüler innerhalb einer Befragung einzelne Items nicht beantworten (*„item-nonresponse“*) (Jekauc, Völkle, Lämmle & Woll, 2012; Schafer & Graham, 2002). Letzteres ist im FOSIS2-Projekt sehr selten, da die Fragebögen innerhalb der Schulstunden angeleitet und im Beisein eines Projektmitarbeiters ausgefüllt und auch auf Vollständigkeit überprüft wurden. Die Item-Nonresponse-Rate liegt für die in dieser Arbeit untersuchten Zielkriterien bei unter 1 %. Im Gegensatz dazu ist die Wave-Nonresponse-Rate im FOSIS2-Projekt deutlich höher. Nach Angaben des Statistischen Bundesamtes lag die durchschnittliche Anzahl an Schülern pro Klasse der Sekundarstufe I im zweiten FOSIS2-Projektjahr, dem Schuljahr 2010/2011, in Berlin bei 22 (Baumann, Schneider, Vollmar & Wolters, 2012). Lägen die FOSIS2-Schulen im Landes-Durchschnitt, müssten die 22 befragten FOSIS2-Klassen in Berlin somit 484 Schüler umfassen. Dieser Orientierungswert entspricht fast der tatsächlichen Maximal-Beteiligung von 479 Schülern zu T2. Angesichts dieser Zahlen erscheint die Gesamtzahl der Fälle ( $n = 570$ ; siehe Tabelle 2), für die Daten von mindestens einer der großen Evaluationsbefragungen vorliegen, sehr hoch. Tatsächlich existieren auch nur für 55 % dieser Fälle ( $n = 314$ ; siehe Tabelle 2) vollständige Datensätze. Für die Analysen zur Implementation werden neben den Daten von T1 und T2 zusätzlich die MSR-Befragungen herangezogen.

Von den Interventionsschülern, die an einer dieser Befragungen teilgenommen haben, sind nur für 47.6 % der Fälle vollständige Datensätze vorhanden (n = 130; siehe Tabelle 3). Anders als bei klinischen Längsschnittstudien, wo beispielsweise durch Mortalität oder auch Interventionsvermeidung in der Regel ein reiner Drop-Out-Prozess, also zunehmender Stichprobenverlust durch Ausscheiden aus der Studie zu beobachten ist (Gollwitzer & Jäger, 2009), gibt es im FOSIS2-Projekt innerhalb der Gruppe von Schülern mit unvollständigen Datensätzen jedes erdenkliche Teilnahme-Muster: z. B. Schüler, für die nur zu T1 oder T3 Daten vorliegen, oder Schüler, die nur an den MSR-Befragungen teilgenommen haben.

Wie lässt sich diese Fluktuation in der Befragungsteilnahme erklären? Der größte Teil der Fluktuation ist vermutlich auf Probleme beim Zusammenführen („*matching*“) der Fälle über die Zeit zurückzuführen. Um die verschiedenen messzeitpunktspezifischen Datensätze eines Schülers aneinanderfügen zu können, wurden die Schüler gebeten, einen individualisierten Code<sup>32</sup> anzugeben, der die Zuordnung der verschiedenen Befragungen zu einer Person erlaubt, ohne Rückschlüsse auf die Identität des Schülers zu ermöglichen. Manche dieser Codes waren unleserlich und einige Schüler gaben unabsichtlich unterschiedliche beziehungsweise falsche Codes an (z. B. statt der ersten zwei Buchstaben des Vornamens der Mutter die Initialen oder statt des eigenen Geburtstags den Monat oder das Geburtsjahr). Da die Codes der Anonymisierung dienten, konnten sie während der Befragung nicht auf ihre Richtigkeit überprüft werden. Mithilfe eines aufwändigen Cleaning-Verfahrens<sup>33</sup> wurden zwar Datensätze von 147 Schülern nachträglich zusammengeführt. Es ist jedoch zu vermuten, dass nicht alle fehlerhaften Codes beim Cleaning identifiziert wurden und somit Daten ein und desselben Schülers fälschlicherweise als Daten von zwei verschiedenen Personen behandelt werden. Dadurch wird die Gesamtheit aller Schüler, die an mindestens einer Befragung teilgenommen haben, überschätzt. Eine andere mögliche Erklärung für eine „Wave-Nonresponse“ wäre, dass Schüler am Tag der Befragung wegen Krankheit oder anderen Gründen (z. B. Schwänzen, familiären Angelegenheiten) nicht anwesend waren. Zudem könnten Schüler auch vereinzelt die Schule oder die Klasse (z. B. aufgrund von Sitzenbleiben oder Umzug) gewechselt haben und somit ganz aus dem Projekt ausgeschie-

---

<sup>32</sup> Der Code setzt sich aus den ersten zwei Buchstaben des Vornamens der Mutter, dem eigenen Geburtstag und den ersten zwei Buchstaben des Vornamens des Vaters zusammen. Bei Unbekanntheit der elterlichen Vornamen wurden die Schüler gebeten, „XX“ einzutragen.

<sup>33</sup> Für die nachträgliche Zuordnung und Korrektur der Codes wurden alle Fragebögen von Schülern, die nicht „matchbar“ waren, von zwei unabhängigen Beurteilern überprüft. Dabei wurden sowohl die Handschrift als auch die demografischen Merkmale hinsichtlich ihrer Ähnlichkeit verglichen. Plädierten beide Beurteiler für eine Übereinstimmung, wurden die Codes entsprechend korrigiert.

den beziehungsweise neu in die Klassen hinzugestoßen sein. Nach Angaben des Statistischen Bundesamtes bleiben durchschnittlich 2 % der Schüler eines Jahrgangs sitzen (Baumann et al., 2012). Die Fluktuation aufgrund von Sitzenbleiben ist also als gering einzuschätzen. Um Neuzugänge aus den Befragungen auszunehmen, wurden bei der zweiten und dritten Befragung nur die Schüler zur Teilnahme aufgefordert, die bereits seit der siebten Klasse auf der Schule waren und somit auch am ersten Projektjahr teilgenommen hatten. Trotzdem lässt sich nicht vollständig ausschließen, dass vereinzelt auch Schüler den Evaluationsfragebogen zu T2 oder T3 ausgefüllt haben, die beispielsweise durch Umzug, Schulwechsel oder Sitzenbleiben erst zum zweiten Schuljahr hinzugekommen sind. Da dieser Anteil an wirklichen Neuzugängen jedoch als gering zu erachten ist, stellt sich die Frage, wie der für Längsschnittuntersuchungen eher untypische Zuwachs an Schülern vom ersten zum zweiten Messzeitpunkt in den Kontrollschulen (siehe Tabelle 2) erklärt werden kann. Dieser lässt sich vor allem mit einer mangelnden Bekanntheit des FOSIS2-Projekts und damit verbundener Schwierigkeiten bei der Organisation der Prä-Befragung in den Kontrollschulen begründen: Die Prä-Befragung fand in allen Schulen innerhalb der ersten zwei Wochen nach Schuljahresbeginn 2009, teilweise sogar an den ersten Schultagen, statt. Da in Berlin der Wechsel auf die weiterführende Schule nach sechs Jahren Grundschule erfolgt, wurden die siebten Klassen zu Schuljahresbeginn völlig neu zusammengesetzt. Die meisten Schulen hatten zum Zeitpunkt der Befragungen noch keine klaren Klassen-, Raum- und Stundenpläne. Für die Umsetzung der Befragung waren daher genaue Absprachen mit den Klassenlehrern nötig, um die Klassen möglichst vollständig bei der Befragung zu erreichen. Genau diese Absprachen und die Organisation der Befragung gestalteten sich in den Kontrollschulen schwieriger als in den Interventionsschulen. Letzteren war das FOSIS2-Projekt bereits seit Ende des Schuljahres 2008/2009 bekannt: Im Rahmen von Einführungsworkshops waren das Projekt vorgestellt, jeder Interventionsschule eine wissenschaftliche Begleitperson zugeordnet, Lehrkräfte als Ansprechpartner für die Projektmitarbeiter benannt und erste Termine für die Befragungen und Workshops vereinbart worden. Im Gegensatz dazu waren die Kontrollschulen erst innerhalb der Sommerferien 2009 durch die Schulleiter der Interventionsschulen rekrutiert worden. Dadurch war der Bekanntheitsgrad des FOSIS2-Projekts bei den Lehrkräften der Kontrollschulen geringer als in den Interventionsschulen und es gab zunächst keine Lehrkräfte als Kontaktpersonen für die Vereinbarung von Befragungs-Terminen. Diese wurden über das Sekretariat der Schulleitung vergeben, ohne dabei Sondertermine der Klassen zu beachten. So konnten drei Klassen einer Kontrollschule bei der Prä-Befragung nur zur Hälfte erfasst werden,



weil die andere Hälfte der Schüler im Rahmen von Kurs-Unterricht an einer Exkursion teilnahm. Zudem verzeichneten die Kontrollschulen nach eigenen Angaben in ihren siebten Klassen innerhalb der ersten vier Schul-Wochen noch Neuzugänge, die ebenfalls bei der Prä-Befragung nicht erreicht werden konnten. Eine Nachbefragung der zu T1 nicht erfassten Schüler war nach Rücksprache mit den Schulleitungen aufgrund des Stundenausfalls durch die Befragungen nicht möglich. Die Termine für die zweite und dritte Evaluationsbefragung konnten mit den Klassenlehrern der Kontrollklassen vereinbart werden, sodass bei den späteren Befragungen ähnliche Erreichbarkeits-Probleme vermieden wurden.

Wie wird mit den fehlenden Werten, speziell mit dem ‚Wave-Nonresponse‘, in dieser Untersuchung verfahren? In der psychologischen Forschung gibt es verschiedene Ansätze zum Umgang mit fehlenden Werten (für eine ausführliche Darstellung der Methoden siehe Lüdtke et al., 2007 und Schafer & Graham, 2002). Das am häufigsten praktizierte Vorgehen ist der sogenannte *fallweise Ausschluss* (Jekauc et al., 2012). Hierbei werden nur solche Personen in die Analysen einbezogen, die an allen Messzeitpunkten teilgenommen haben (Jekauc et al., 2012; Lüdtke et al., 2007; Schafer & Graham, 2002). Der Vorteil dieses Verfahrens liegt in der besonders leichten Umsetzung: Alle unvollständigen Fälle werden einfach aus den Analysen ausgeschlossen. Mögliche Nachteile können die Reduktion der Stichprobengröße und damit verbundene verzerrte Parameterschätzungen sein (Lüdtke et al., 2007). Speziell für den Umgang mit fehlenden Werten durch ‚Wave-Non-Response‘ empfehlen Schafer und Graham (2002), wenn möglich, *Maximum-Likelihood-Verfahren* (ML) einzusetzen:

*„Because repeated measures on an individual tend to be correlated, we recommend procedures that use all available data for each participant, because missing information can then be partially recovered from earlier or later waves. Longitudinal modeling by ML can be a highly efficient way to use all available data“ (S. 150).*

Der ML-Ansatz zählt nach Lüdtke et al. (2007) zu den modellbasierten Verfahren zum Umgang mit fehlenden Werten. Diese zeichnen sich dadurch aus, dass sie alle vorhandenen Daten berücksichtigen können und nicht auf vollständige Datensätze angewiesen sind. Welcher Ansatz zum Umgang mit fehlenden Werten verfolgt wird, ist abhängig von der gewählten Auswertungsmethode und auch von der konkreten Beschaffenheit der fehlenden Werte. Für den fallweisen Ausschluss sollten beispielsweise vollständig zufällig fehlende Werte vorliegen (*missing completely at random*, MCAR), für das ML-basierte Vorgehen

reichen hingegen zufällig fehlende Werte (*missing at random*‘, MAR) aus.<sup>34</sup> Welche Form der *missing Data*‘ vorliegt, lässt sich nicht abschließend ermitteln (Lüdtke et al., 2007). Um die fehlenden Werte genauer zu analysieren, sollte aber beispielsweise geprüft werden, ob sich die Personen mit fehlenden Werten systematisch von Personen mit vollständigen Datensätzen unterscheiden (Jekauc et al., 2012). Solche Analysen werden nachfolgend im Abschnitt „Wave-Nonresponse-Analysen“ beschrieben. Dabei werden zunächst Drop-Out-Analysen vorgenommen, das heißt, die Gruppe der Schüler, die an T1 teilgenommen hat und anschließend irgendwann aus dem Projekt ausgeschieden ist (*Drop-Out-Stichprobe*‘), wird in den demografischen Variablen und den Zielkriterien zu T1 mit den Schülern mit vollständigen Datensätzen (*Completer-Stichprobe*‘) verglichen. Wie oben beschrieben ist im FOSIS2-Projekt jedoch nicht nur ein Drop-Out-Prozess zu beobachten, sondern die Beteiligung an den Befragungen variiert über alle Messzeitpunkte. Deshalb werden im Anschluss an die Drop-Out-Analysen zusätzlich alle Schüler mit unvollständigen Datensätzen (*Incompleter-Stichprobe*‘) mit der *Completer-Stichprobe*‘ hinsichtlich systematischer Unterschiede in den demografischen Variablen und den Zielkriterien zu allen Messzeitpunkten verglichen. Darüber hinaus wird mithilfe des „MCAR-Tests“ nach Little (1988) überprüft, ob die fehlenden Werte von anderen beobachteten Variablen abhängen. Bei Nicht-Signifikanz ist dieser Test ein Hinweis auf die Gültigkeit der MCAR-Annahme (Lüdtke et al., 2007).<sup>35</sup> Da für die Beantwortung der Fragestellungen dieser Arbeit Daten von unterschiedlichen Befragungen herangezogen werden, erfolgen die Wave-Nonresponse-Analysen getrennt für jede Fragestellung.

---

<sup>34</sup> Methodisch werden drei Formen von *missing data*‘ unterschieden: Vollständig zufällig fehlende Werte (*missing completely at random*‘, MCAR), zufällig fehlende Werte (*missing at random*‘, MAR) und nicht zufällig fehlende Werte (*missing not at random*‘, MNAR). Bei den MCAR hängt „das Auftreten von fehlenden Werten weder von der Ausprägung der Variablen selbst, noch von der Ausprägung anderer Variablen des Datensatzes ab“ (Lüdtke et al, 2007, S. 104). Die fehlenden Werte können somit als Zufallsstichprobe betrachtet werden. Eine etwas schwächere Annahme gilt für fehlende Daten, die als MAR bezeichnet werden. Hier sollten die fehlenden Werte zwar weiterhin nicht von der Ausprägung der Variablen selbst abhängen, dürfen aber systematische Zusammenhänge mit anderen beobachteten Variablen aufweisen. MNAR trifft auf alle Arten fehlender Werte zu, die nicht als MAR bezeichnet werden, wo also neben Zusammenhängen mit anderen beobachteten Variablen auch Zusammenhänge zwischen den fehlenden Werten und der Ausprägung auf den fehlenden Variablen selbst bestehen (Jekauc et al., 2012; Lüdtke et al., 2007).

<sup>35</sup> Zu beachten ist, dass durch Little’s MCAR-Test nicht geprüft werden kann, ob die fehlenden Werte mit der Ausprägung der Variablen selbst zusammenhängen, und dass somit die MCAR-Bedingung nicht vollständig verifizierbar ist (Lüdtke et al., 2007).

### ***Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3)***

Für die Daten zur Beantwortung der ersten Fragestellung (T1, T2, T3) zeigen sich zum ersten Messzeitpunkt keine systematischen Unterschiede in den demografischen Variablen und den Zielkriterien zwischen der ‚Drop-Out-Stichprobe‘ und den Schülern mit vollständigen Datensätzen (siehe Anhang A1, Tabelle A1). Auch die gesamte ‚Incompleter-Stichprobe‘ unterscheidet sich zu T1 nicht systematisch von den Schülern mit vollständigen Datensätzen (siehe Anhang A1, Tabelle A2). Hingegen unterscheiden sich ‚Completer‘ und ‚Incompleter‘ zum zweiten Messzeitpunkt (T2) systematisch im Alter, der schulischen Selbstwirksamkeitserwartung und dem Verlusterleben: Die Schüler mit unvollständigen Datensätzen sind signifikant älter und weisen eine bedeutsam geringere schulische Selbstwirksamkeit sowie ein höheres Verlusterleben auf (siehe Anhang A1, Tabelle A3). Mithilfe einer binären logistischen Regression wurden die Variablen, für die es signifikante Unterschiede gab, simultan im Hinblick auf ihre Vorhersagekraft zur Gruppenzugehörigkeit (‚Completer‘ versus ‚Incompleter‘) getestet. Das Gesamtmodell wird signifikant ( $\chi^2 = 27.892$ ;  $df = 3$ ;  $p < 0.001$ ). Von den eingeschlossenen Variablen erweist sich jedoch im simultanen Test nur das Alter als bedeutsam für die Vorhersage der Gruppenzugehörigkeit (Alter zu T2:  $B = -0.732$ ;  $SE = 0.164$ ;  $p < 0.001$ ; Schulische Selbstwirksamkeitserwartung zu T2:  $B = 0.186$ ;  $SE = 0.207$ ;  $p > 0.05$ ; Verlusterleben zu T2:  $B = 0.897$ ;  $SE = 0.156$ ;  $p > 0.05$ ). Zu T3 unterscheiden sich Schüler mit vollständigen und unvollständigen Datensätzen ebenfalls im Alter und der schulischen Selbstwirksamkeitserwartung und zusätzlich in der erlebten Binnendifferenzierung (siehe Anhang A1, Tabelle A4). Die drei signifikanten Variablen werden wiederum simultan in eine binäre logistische Regression aufgenommen, um ihre Vorhersagekraft auf die Gruppenzugehörigkeit (‚Completer‘ versus ‚Incompleter‘) zu prüfen. Wieder wird das Gesamtmodell signifikant ( $\chi^2 = 20.125$ ;  $df = 3$ ;  $p < 0.001$ ) und auch hier erweist sich nur das Alter als bedeutsam für die Vorhersage der Gruppenzugehörigkeit (Alter zu T3:  $B = -0.576$ ;  $SE = 0.175$ ;  $p < 0.001$ ; Schulische Selbstwirksamkeitserwartung zu T3:  $B = 0.284$ ;  $SE = 0.184$ ;  $p > 0.05$ ; Binnendifferenzierung zu T3:  $B = 0.311$ ;  $SE = 0.161$ ;  $p > 0.05$ ). Somit ist festzuhalten, dass die Zugehörigkeit zur Gruppe von Schülern mit vollständigen und unvollständigen Datensätzen nicht durch die Zielkriterien, jedoch durch das Alter vorhergesagt werden kann. Sowohl zu T2 als auch zu T3 waren ältere Schüler mit höherer Wahrscheinlichkeit in der Gruppe mit unvollständigen Datensätzen. Dies mag daran liegen, dass ältere Schüler eher dem Unterricht fernbleiben als jüngere Schüler (Wagner, 2007). Auch die bereits oben erwähnten Wieder-

holer könnten für das bedeutsam höhere Alter der Schüler mit unvollständigen Datensätzen verantwortlich sein. Der MCAR-Test nach Little (1988) ergab, dass die fehlenden Werte für Daten der drei Evaluationsbefragungen unter Berücksichtigung aller demografischen Variablen und aller erfassten Kriterien nicht systematisch mit den Ausprägungen der Schüler auf anderen Variablen zusammenhängen (Little's MCAR-Test für T1-T2-T3:  $\chi^2 = 5361.116$ ;  $df = 5616$ ;  $p > 0.05$ ).

Zusammenfassend kann die Zugehörigkeit zur Gruppe von Schülern mit vollständigen beziehungsweise unvollständigen Werten sowohl zu T2 als auch zu T3 signifikant durch das Alter vorhergesagt werden. Unter Hinzunahme aller beobachteten Variablen gibt es nach dem MCAR-Test jedoch keine Hinweise auf systematische Zusammenhänge zwischen den fehlenden Werten und den Ausprägungen auf anderen Variablen.

***Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2)***

Für die Daten zur Beantwortung der Frage nach der Wirksamkeit des FOSIS2-Programms in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3, T2, nur Interventionsschüler) zeigen sich zum ersten Messzeitpunkt ebenfalls keine systematischen Unterschiede in den demografischen Variablen und den Zielkriterien zwischen der ‚Drop-Out-Stichprobe‘ und den Interventionsschülern mit vollständigen Datensätzen (siehe Anhang A2, Tabelle A5). Auch der Vergleich der ‚Incompleter-‘ mit der ‚Completer-Stichprobe‘ zeigt keine systematischen Unterschiede zu T1: Die Interventionsschüler mit unvollständigen Daten zu T1, MSR1 bis MSR3 und T2 unterscheiden sich nicht bedeutsam von den Interventionsschülern mit vollständigen Datensätzen zu allen fünf Messzeitpunkten (siehe Anhang A2, Tabelle A6). Auch zu T2 weisen die Interventionsschüler mit unvollständigen und vollständigen Datensätzen keine systematischen Unterschiede in den demografischen Variablen und den Zielkriterien auf (siehe Anhang A2, Tabelle A7). Der MCAR-Test nach Little (1988) ergab, dass die fehlenden Werte für Daten der MSR1 bis MSR3 und der ersten beiden Evaluationsbefragungen unter Berücksichtigung aller demografischer Variablen, aller erfasster Kriterien und der Implementationswahrnehmung nicht systematisch mit den Ausprägungen der Interventionsschüler auf anderen Variablen zusammenhängen (Little's MCAR-Test für T1-MSR1-3-T2:  $\chi^2 = 5051.369$ ;  $df = 5238$ ;  $p > 0.05$ ).

Über alle Vergleiche und Tests hinweg zeigen sich folglich weder zu T1 noch zu T2 systematische Unterschiede zwischen den Interventionsschülern mit vollständigen Datensät-

zen (zu T1, MSR1 bis MSR3 und T2) und den Schülern mit unvollständigen Datensätzen. Auch nach dem MCAR-Test finden sich keine Zusammenhänge zwischen den fehlenden Werten und der Ausprägung auf anderen beobachteten Variablen.

### ***Wahl der Analysestichproben***

Die statistischen Analysen der vorliegenden Untersuchung werden mit dem Verfahren der Mehrebenenanalyse mittels HLM 6.08 durchgeführt (in Abschnitt 6.5 und 6.5.3 wird das Vorgehen ausführlich beschrieben). Liegen mehr als zwei Messzeitpunkte vor, ist im Rahmen von Mehrebenenanalysen die Modellierung der Zeit auf einer eigenen Ebene möglich. Hierbei kann das beschriebene modellbasierte ML-Verfahren zum Umgang mit fehlenden Werten realisiert werden (Hox, 2010). Für die Beantwortung der Frage nach der Wirksamkeit des FOSIS2-Programms, wird eine solche Zeit-Ebene modelliert (siehe Abschnitt 6.5) und das ML-Verfahren eingesetzt. In die Analysen und Parameterschätzungen fließen folglich Informationen von allen vorhandenen Fällen ein ( $n = 570$ ; siehe Tabelle 2). Beim ML-Verfahren muss nicht die MCAR-Annahme, sondern nur die MAR-Bedingung erfüllt sein (Jekauc et al., 2012). Zusammenhänge zwischen den fehlenden Werten und anderen beobachteten Variablen, wie sie beispielsweise für das Alter gefunden wurden, sind also erlaubt.

Der Einfluss der wahrgenommenen Implementation auf die Programm-Wirksamkeit soll beispielhaft für das erste Projektjahr untersucht werden. Entsprechend werden nur die erste und zweite Evaluationsbefragung herangezogen und es kann keine eigene Ebene der Zeit modelliert werden (siehe auch Abschnitt 6.5.4). Bei der Parameterschätzung im Rahmen solcher Modelle zieht das Programm HLM 6.08 nur Informationen von Schülern mit vollständigen Datensätzen heran ( $n = 130$ ; siehe Tabelle 3). Für diesen fallweisen Ausschluss sollte die beschriebene MCAR-Bedingung erfüllt sein (Jekauc et al., 2012). Da sich bei den Wave-Non-Response-Analysen keinerlei systematische Zusammenhänge zwischen den Interventionsschülern mit vollständigen Datensätzen (zu T1, MSR1 bis MSR3 und T2) und den Schülern mit unvollständigen Datensätzen zeigen und Little's MCAR-Test nicht signifikant ist, scheint die MCAR-Annahme für die zweite Analysestichprobe haltbar zu sein. Zudem handelt es sich bei den Analysen um eine rein explorative Betrachtung des möglichen Zusammenhangs von Implementationswahrnehmung und der Programmwirksamkeit. Die Ergebnisse werden als Pilot-Ergebnisse gewertet, die erste Hinweise auf mögliche Zusammenhänge zwischen der modulspezifischen Implementationswahrnehmung der Interventionsschüler und den Kriterien liefern sollen. Da die Analysen explorativ erfolgen, wird

für die Analysestichprobe zur Beantwortung der Frage nach der Wirksamkeit in Anhängigkeit von der wahrgenommenen Implementation die mit dem fallweisen Ausschluss verbundene Reduktion der Stichprobe in Kauf genommen.

### **6.3.3 Beschreibung der Analysestichproben**

In Tabelle 4 sind die demografischen Merkmale der Analysestichprobe für die Evaluation der Wirksamkeit des FOSIS2-Programms dargestellt. Durchschnittlich sind 25 Schüler in den 11 Interventionsklassen und 27 in den 11 Kontrollklassen. Von den 271 Interventionsschülern, die an mindestens einer der drei Evaluationsbefragungen teilgenommen haben, sind 47.6 % Mädchen und 51.3 % Jungen. Die 299 Schüler der Kontrollgruppe sind zu 53.9 % weiblich und zu 43.1 % männlich. Die Geschlechterverteilungen beider Gruppen unterscheiden sich nicht signifikant. Im Durchschnitt sind die Interventionsschüler zum ersten Messzeitpunkt 12 Jahre und 4 Monate und die Kontrollschüler 12 Jahre und 5 Monate alt. Der Altersunterschied in den beiden Gruppen ist ebenfalls nicht signifikant. Laut eigenen Angaben sind 69.4 % der Interventionsschüler deutsche Staatsbürger, 18.8 % berichten, keine deutsche Staatsbürgerschaft zu haben. In der Kontrollgruppe liegt das Verhältnis bei 82.3 % mit deutscher und 15.0 % mit einer anderen Staatsbürgerschaft. Auch hier unterscheiden sich die Verteilungen in der Interventions- und der Kontrollgruppe nicht signifikant (siehe Tabelle 5).

**Tabelle 4:** Analyse-Stichprobe für die Evaluation der Wirksamkeit (T1-T2-T3)

	Interventions- gruppe	Kontrollgruppe	Chi-Quadrat-/ T-Test	Gesamt
<b>Analyse-Stichprobe:</b>				
Schüler	271	299	--	570
Klassen (Schulen)	11 (3)	11 (3)		22 (6)
durchschnittliche Anzahl d. Schüler pro Klasse	24.6	27.2		25.9
<b>Geschlecht:</b>				
Weiblich	129 (47.6%)	161 (53.9 %)	n.s.	290 (50.9%)
Männlich	139 (51.3%)	129 (43.1%)		268 (47%)
keine Angabe	3 (1.1%)	9 (3.0%)		12 (2.1%)
<b>Alter:</b>				
M (Jahre, Monate)	12;4	12;5	n.s.	
SD (Monate)	8	8		
<b>Nationalität:</b>				
Deutsch	188 (69.4%)	246 (82.3%)	n.s.	434 (76.0%)
andere Nationalität	51 (18.8%)	45 (15.0%)		96 (17.9%)
Keine Angabe	32 (11.8%)	8 (2.7%)		40 (6.1%)

Anmerkungen: M = Mittelwert; SD = Standardabweichung; n.s. = nicht signifikant

In Tabelle 5 sind die demografischen Merkmale der Analytestichprobe für die Evaluation der Wirksamkeit des FOSIS2-Programms in Abhängigkeit von der wahrgenommenen Implementation seitens der Interventionsschüler dargestellt. Wie oben beschrieben konnten zur MSR2 nur zehn Interventionsklassen befragt werden. In diesen sind durchschnittlich 13 Schüler, die an T1, MSR1 bis MSR3 und T2 teilgenommen haben. Von den insgesamt 130 Schülern mit vollständigen Datensätzen sind 50.8 % Mädchen und 49.2 % Jungen. Das mittlere Alter der Schüler beträgt zum ersten Messzeitpunkt 12 Jahre und vier Monate. Insgesamt verfügen 70.8 % der Schüler laut eigenen Angaben über eine deutsche und 16.9 % über eine andere Staatsbürgerschaft.

**Tabelle 5:** Analyse-Stichprobe für die Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation der Interventionsschüler (T1, MSR1-3, T2)

Interventionsgruppe	
<b>Analyse-Stichprobe:</b>	
Schüler	130
Klassen (Schulen)	10 (3)
durchschnittliche Anzahl d. Schüler pro Klasse	13
<b>Geschlecht:</b>	
weiblich	66 (50.8%)
männlich	64 (49.2%)
Keine Angabe	--
<b>Alter:</b>	
M (Jahre; Monate)	12;4
SD (Monate)	8
<b>Nationalität:</b>	
Deutsch	92 (70.8%)
andere Nationalität	22 (16.9%)
Keine Angabe	16 (12.3%)

Anmerkungen: M = Mittelwert; SD = Standardabweichung

## 6.4 Erhebungsinstrumente

Den Analysen zur Beantwortung der Fragestellungen dieser Arbeit liegen Daten zu den motivational relevanten Konstrukten der standardisierten Prä-, Zwischen- und Post-Befragungen der Schüler sowie der Implementationsindikatoren der standardisierten modulspezifischen Schüler-Rückmeldungen des ersten Projektjahres zugrunde. Nachfolgend werden die Instrumente zur Erfassung der motivationalen Zielkriterien und der von den Schülern wahrgenommenen Implementation sowie ihre Reliabilitäten und Trennschärfen der jeweiligen Skalen beschrieben. Eine Auflistung der einzelnen Items sowie eine Übersicht über weitere Kennwerte der Skalen finden sich im Skalenhandbuch im Anhang B.

### 6.4.1 Skalen der Prä-, Zwischen- und Postbefragung

Bereits in Kapitel 5 wurden die Zielkriterien ausführlich geschildert, die für die Evaluation des FOSIS2-Programms herangezogen und im Rahmen der Prä-, Zwischen- und Postbefragung (T1, T2, T3) der Schüler erhoben wurden. In motivationaler Hinsicht sollen sich die Maßnahmen des FOSIS2-Projekts auf die schulische Selbstwirksamkeitserwartung, die Prüfungsängstlichkeit, die schulbezogene Stresswahrnehmung und die wahrgenommene Binnendifferenzierung auswirken. Die Skalen zur Erfassung dieser Konstrukte wurden zu allen drei Befragungen unverändert eingesetzt und konnten mit einer vierstufigen Likert-



Skala von „trifft nicht zu“ (1), „trifft kaum zu“ (2), „trifft eher zu“ (3) und „trifft genau zu“ (4) beantwortet werden.

### ***Schulische Selbstwirksamkeitserwartung***

Die hier verwendeten Items zur Erfassung der schulischen Selbstwirksamkeitserwartung basieren auf der von Jerusalem und Satow (1999) entwickelten Skala „Schulbezogene Selbstwirksamkeitserwartung“. Mithilfe von fünf Items misst die Skala die Einschätzung der eigenen Kompetenzen im Umgang mit schulischen Anforderungen, also die schulspezifische Kompetenzerwartung der Schüler. Ein Beispielitem der Skala lautet „Ich kann auch die schwierigen Aufgaben im Unterricht lösen, wenn ich mich anstrengte“. Die interne Konsistenz (Cronbachs  $\alpha$ ) ist in allen drei Befragungen zufriedenstellend<sup>36</sup> (T1 :  $\alpha = .72$ ; T2:  $\alpha = .75$ ; T3:  $\alpha = .77$ ). Auch die korrigierten Trennschärfen der Items, die zwischen  $r_{itc} = .44$  und  $r_{itc} = .56$  liegen, sind als annehmbar zu bewerten.<sup>37</sup>

### ***Prüfungsängstlichkeit***

Die hier verwendete Skala zur Erfassung der Prüfungsängstlichkeit wurde in Anlehnung an die Besorgtheits-Komponente des deutschen TAI von Hodapp et al. (Hodapp et al., 1982) entwickelt und bereits in Vorgängerprojekten erprobt (siehe z. B. Jerusalem, 2008; Röder, 2008). Mit der aus fünf Items bestehenden Skala wird die Besorgtheit als Bestandteil der Ängstlichkeit im Sinne einer überdauernden Persönlichkeitseigenschaft erhoben. Nach Schwarzer (2000) ist die Besorgtheit das wichtigste Kernmerkmal der Prüfungsängstlichkeit. Ein Beispielitem lautet „Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, bin ich besorgt, dass etwas schief laufen könnte“. Die interne Konsistenz ist mit Werten von  $\alpha = .82$  zu T1,  $\alpha = .85$  zu T2 und  $\alpha = .87$  zu T3 als sehr zufriedenstellend zu bewerten und auch die korrigierten Trennschärfen liegen in einem hohen Bereich und variieren zwischen  $r_{itc} = .54$  und  $r_{itc} = .76$ .

---

<sup>36</sup> Rost (2013) gibt folgende Empfehlung für die Bewertung der internen Konsistenz: Bei der Individualdiagnostik (z. B. bei Intelligenztests und anderen Leistungstests) ist eine Reliabilität ab .75 als gut, eine Reliabilität ab .80 als sehr gut und eine Reliabilität ab .85 als exzellent zu bewerten. „Bei Forschungsprojekten, in denen man nur an einem Gruppenvergleich interessiert ist, kann die Reliabilität deutlich niedriger liegen. Der Zuverlässigkeitskoeffizient sollte [...] aber mindestens .55 betragen.“ (Rost, 2013, S. 179). Im Fokus der vorliegenden Arbeit stehen Gruppenvergleiche wie der Vergleich zwischen Interventions- und Kontrollgruppe. Deshalb wird eine interne Konsistenz ab .55 als zufriedenstellend bewertet.

<sup>37</sup> Nach Bühner (2011) sind Trennschärfen, die kleiner als .30 ausfallen, als niedrig, Trennschärfen zwischen .30 und .50 als mittel und Trennschärfen größer als .50 als hoch zu bewerten.

### ***Schulbezogenes Stresserleben – Herausforderung, Bedrohung und Verlust***

In Anlehnung an die im transaktionalen Stressmodell von Lazarus und Folkman (1984) operationalisierten Stresseinschätzungen Herausforderung, Bedrohung und Verlust wird das schulbezogene Stresserleben im FOSIS2-Projekt mittels dreier gleichnamiger Skalen erfasst. Die Skalen bestehen jeweils aus drei Items, die ursprünglich von Jerusalem (1990) entwickelt und im Rahmen von FOSIS2-Vorgängerprojekten an den schulischen Kontext angepasst wurden (Jerusalem, 2005a, 2008). Mithilfe der Skala „wahrgenommene schulbezogene Herausforderung“ soll das Ausmaß gemessen werden, in welchem sich ein Schüler durch die Schule positiv herausgefordert fühlt. Ein Beispielitem der Skala lautet „Die Schule ist interessant, weil ich täglich neu herausgefordert werde“. Sowohl die interne Konsistenz der Skala als auch die korrigierten Trennschärfen sind als zufriedenstellend zu bewerten (T1:  $\alpha = .69$ ; T2:  $\alpha = .71$ ; T3:  $\alpha = .79$ ;  $r_{\text{itc}} = .56$  bis  $r_{\text{itc}} = .68$ ). Die Skala „wahrgenommene schulbezogene Bedrohung“ erfasst die Befürchtung, nicht über ausreichend Bewältigungskompetenzen zu verfügen, um den schulischen Anforderungen gerecht zu werden. Eines der drei Items dieser Skala heißt „Ich befürchte, dass ich mit meinen Problemen in der Schule nicht zurechtkomme“. Mit einem  $\alpha$  von .54 ist die interne Konsistenz zu T1 nach Rost (2013) noch knapp akzeptabel. Zu T2 und T3 ist die Reliabilität mit  $\alpha = .71$  und  $\alpha = .77$  deutlich besser. Die korrigierten Trennschärfen sind ebenfalls als annehmbar zu bewerten ( $r_{\text{itc}} = .33$  bis  $r_{\text{itc}} = .68$ ). Anders als die Herausforderungs- und Bedrohungseinschätzungen, bei denen es sich um künftige Erwartungen beziehungsweise Befürchtungen handelt, zielt die Skala „wahrgenommener schulbezogener Verlust“ auf einen bereits eingetretenen subjektiv erlebten Kontrollverlust in der Schule ab. Eines der drei Items lautet „Meine augenblickliche schulische Lage ist hoffnungslos“. Reliabilität und korrigierte Trennschärfen fallen zu allen drei Messzeitpunkten akzeptabel aus (T1:  $\alpha = .60$ ; T2:  $\alpha = .75$ ; T3:  $\alpha = .82$ ;  $r_{\text{itc}} = .32$  bis  $r_{\text{itc}} = .68$ ).

### ***Wahrgenommene Binnendifferenzierung***

Die ebenfalls nur aus drei Items bestehende Skala „Binnendifferenzierung“ wurde in Anlehnung an den „Individualized Classroom Environment Questionnaire (ICEQ)“ von Fraser (1982) entwickelt. Aus der Perspektive der Schüler erfassen die Items, inwiefern die Schüler binnendifferenzierende beziehungsweise individualisierende Maßnahmen im Unterricht erleben, das heißt, ob sich die Lehrkräfte sowohl bei den Aufgabenstellungen als auch bei Leistungsrückmeldungen am individuellen Leistungsstand der Schüler orientieren. Ein Beispielitem lautet „Die weniger guten Schüler erhalten leichtere Aufgaben als die

guten Schüler“. Mit Werten von  $\alpha = .56$  zu T1,  $\alpha = .46$  zu T2 und  $\alpha = .54$  zu T3 ist die interne Konsistenz der Skala nicht zufriedenstellend und nur knapp akzeptabel. Auch die korrigierten Trennschärfen liegen niedriger als bei den anderen Skalen und variieren zwischen  $r_{itc} = .19$  und  $r_{itc} = .43$ . Bei der Interpretation der Ergebnisse muss die als eher gering einzuschätzende psychometrische Qualität der Skala unbedingt berücksichtigt werden.

#### **6.4.2 Skalen der modulspezifischen Rückmeldungen**

Gerade bei Interventionen, die durch Mediatoren vermittelt werden, ist die Erfassung der Programmumsetzung unerlässlich, da diese hinsichtlich Ausmaß und Genauigkeit zwischen den Mediatoren variieren kann (Durlak & DuPre, 2008; Scheirer et al., 1995). Am Ende einer jeden Implementationsphase wurden die Schüler und Lehrer der Interventionschulen deshalb mittels standardisierter sogenannter modulspezifischer Rückmeldungen zum Implementationsgeschehen befragt. In der Arbeit werden die Analysen zur Wirksamkeit des FOSIS2-Programms in Abhängigkeit von der Implementation für das erste Projektjahr durchgeführt und hierfür die modulspezifischen Rückmeldungen der Schüler herangezogen. Etwa im Abstand von drei Monaten wurden die Interventionsschüler nach dem erlebten Ausmaß der Umsetzung der FOSIS2-Maßnahmen gefragt. Für jedes Fach, das von einer Lehrkraft unterrichtet wurde, die am Projekt teilgenommen hat, sollten die Schüler einschätzen, wie häufig sie in den letzten Wochen die einzelnen Förderstrategien der Module Klassenklima, Transparenz und Lern- und Leistungsraum im Unterricht wahrgenommen haben. Mithilfe der Handreichungen, in denen die Förderstrategien der FOSIS2-Module festgehalten sind, wurden also modulspezifische Skalen entwickelt, deren Items die Kernprinzipien der einzelnen FOSIS2-Module erfassen. Die Items sollten dann jeweils fachspezifisch beantwortet werden. Die Schüler mussten folglich jedes Item für verschiedene nebeneinander in Spalten aufgeführte Fächer einschätzen. Zum besseren Verständnis der Fragebogenkonstruktion zeigt Abbildung 8 einen Ausschnitt der ersten modulspezifischen Rückmeldung mit Items zur Erfassung der Umsetzung von Förderprinzipien des Moduls Klassenklima. Die erlebte Häufigkeit der Förderstrategien konnten die Schüler mit einem vierstufigen Antwortformat von „nie“ (1), „selten“ (2), „oft“ (3), „bei jeder Gelegenheit“ (4) beurteilen. Nachfolgend werden die Items der modulspezifischen Skalen und deren Reliabilität sowie die korrigierten Trennschärfen der Items dargestellt. Die entsprechenden Kennwerte basieren auf den Skalenwerten des Fachs Mathematik. Da die Items zur Implementation der Module für das Fach Mathematik von allen Interventionsschülern beantwortet wurden, eignet sich dieses Fach am ehesten für die Beurteilung der psycho-

metrischen Qualität der Skalen. Für die späteren Analysen zur Wirksamkeit in Abhängigkeit von der Implementation wurden die modul- und fachspezifischen Skalenwerte zu einem immer noch modulspezifischen, jedoch fachübergreifenden Implementationsindikator aggregiert.<sup>38</sup>


L		Mein Kennwort																n
 <b>Unterrichtsbegleitende Fragen für Schülerinnen und Schüler</b>		Die ersten zwei Buchstaben des Vornamens Deiner Mutter				Dein eigener Geburtstag (wenn Du z.B. am 05.02.1998 geboren bist, trage hier 05 ein, also nur den Tag)				Die ersten zwei Buchstaben des Vornamens Deines Vaters								
		<input type="text"/> <input type="text"/>				<input type="text"/> <input type="text"/>				<input type="text"/> <input type="text"/>								
<b>Liebe Schülerin, lieber Schüler, wie hast Du den Unterricht in den <i>letzten Wochen</i> erlebt?</b>		in Mathe				in Deutsch				in Englisch				in Französisch				
		1 = nie      2 = selten      3 = oft      4 = bei jeder Gelegenheit																
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
Wir haben über das Klassenklima gesprochen.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Wir haben Klassenregeln erstellt.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Gemeinsam mit dem Lehrer haben wir darauf geachtet, dass die Klassenregeln eingehalten werden.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Unser Lehrer hat nach unserer Meinung zum Unterricht gefragt.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Wenn unser Lehrer nach unserer Meinung gefragt hat, war er auch bereit etwas zu verändern.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Abbildung 8: Ausschnitt aus der ersten modulspezifischen Rückmeldung zum Klassenklima

### Implementationscheck – Klassenklima

Die aus fünf Items bestehende Skala zur Erfassung der Implementation der Förderstrategien des Moduls Klassenklima ist in Abbildung 8 vollständig dargestellt. Die Schüler werden gefragt, ob das Klassenklima im Unterricht thematisiert, Klassenregeln aufgestellt und auf deren Einhaltung geachtet wurde. Auch die Möglichkeit des Feedbacks zum Unterricht wird thematisiert. Sowohl die interne Konsistenz der Skala als auch die korrigierten Trennschärfen sind als zufriedenstellend zu bewerten (MSR1 – Implementationscheck Klassenklima:  $\alpha = .79$ ;  $r_{itc} = .46$  bis  $r_{itc} = .65$ ).

### Implementationscheck – Transparenz

Mithilfe von fünf Implementations-Checkitems zum Modul Transparenz wird geprüft, ob die Lehrkräfte die im Modulworkshop angesprochenen Bereiche Transparenz der Anforderungen und Transparenz der Bewertungskriterien auch berücksichtigt und Themen einer Klassenarbeit, Vorbereitungsmöglichkeiten, Art der Aufgaben und die Punktevergabe sowie die Bedeutung der Klassenarbeit für die Gesamtnote auch wirklich transparent gestaltet haben. Ein Beispielitem der Skala lautet „In den letzten Wochen hat uns unser Lehrer vor Tests/Klassenarbeiten ganz genau mitgeteilt, welche Themen und Anforderungen dran kommen“. Auch hier liegen die interne Konsistenz der Skala und die korrigierten Trennschärfen

<sup>38</sup> Zur Beibehaltung des Antwortformats 1 bis 4 wurden die aggregierten Implementationsindikatoren an der Anzahl der erfragten Fächer relativiert.

schärfen in einem zufriedenstellenden Bereich (MSR2 – Implementationscheck Transparenz:  $\alpha = .77$ ;  $r_{itc} = .50$  bis  $r_{itc} = .57$ ).

### ***Implementationscheck – Lern- und Leistungsraum***

Die Items der Skala zur Erfassung der Implementation der Maßnahme Lern- und Leistungsraum bilden die Kernprinzipien dieser Intervention ab. Gefragt wird, ob die Schüler die Möglichkeit hatten, üben und lernen zu können, ohne währenddessen Noten zu erhalten, ob diese notenfreien Übungszeiten für die Schüler erkennbar waren und ob der Lehrer dann auch tatsächlich keine Noten vergeben hat. Zudem werden der Einsatz der individuellen Bezugsnorm der Lehrkraft und die Ermutigung im Umgang mit Fehlern thematisiert. Eines der fünf Items der Skala lautet „In den letzten Wochen hatten wir die Möglichkeit, lernen und üben zu können, ohne dass es dafür Noten gab“. Auch bei dieser Skala fallen die Reliabilität und die korrigierten Trennschärfen zufriedenstellend aus (MSR2 – Implementationscheck Lern- und Leistungsraum:  $\alpha = .74$ ;  $r_{itc} = .42$  bis  $r_{itc} = .54$ ).

## **6.5 Methodisches Vorgehen: Mehrebenenanalysen**

Zur Prüfung der Fragestellungen wurden die Daten mittels *hierarchischer linearer Modelle* (Eid, Gollwitzer & Schmitt, 2013) oder auch sogenannter *Mehrebenenanalysen* (Ditton, 1998) beziehungsweise *Mehrebenenmodellierungen mit Zufallskoeffizienten* (Nezlek, Schröder-Abé & Schütz, 2006) mit dem Programm HLM 6.08<sup>39</sup> (Raudenbush, Bryk & Congdon, 2009) auf der Basis von SPSS(PASW)18-Datensätzen<sup>40</sup> analysiert. Nachfolgend werden zunächst die Gründe für die Wahl des Verfahrens beschrieben. Im Anschluss erfolgt eine allgemeine Darstellung möglicher Mehrebenenmodelle zur Veränderungsmessung. Darauf aufbauend werden dann die in dieser Studie durchgeführten Modelle erläutert. Eine ausführliche Schilderung der Voraussetzungen des Verfahrens, wie die substantielle Varianz zwischen den Gruppen, die Intraklassen-Korrelation (ICC), Anforderungen an die Stichprobengröße und Homogenitäts- beziehungsweise Verteilungsvoraussetzungen, findet sich im Anhang C1. Auch das Vorgehen bei der Modellierung in dieser Studie, spricht die schrittweise Modellentwicklung, die verwendeten Kriterien zur Beurteilung der

---

<sup>39</sup> Je nach Autoren und Softwareprogrammen werden für dieselben Modelle unterschiedliche Begrifflichkeiten und Symbole verwendet Garson (2013b). Da zur Analyse der Berechnungen die Software HLM6.08 (Hierarchical Linear Modeling 6.08) eingesetzt wurde, werden in der vorliegenden Arbeit die im Programm enthaltenen Bezeichnungen und Formeln beibehalten (siehe auch Raudenbush, Bryk, Cheong, Congdon & Du Toit, 2004).

<sup>40</sup> Die Daten wurden mit dem für statistische Analysen üblichen Programm SPSS/PASW18 (Statistical Package for the Social Sciences/Predictive Analysis SoftWare) aufbereitet.

Modellgüte, das gewählte Schätzverfahren und die eingesetzten Maßnahmen zur Zentrierung werden im Anhang C2 detailliert begründet.

### 6.5.1 Gründe für die Anwendung von Mehrebenenanalysen

Mehrebenenmodelle beugen Fehlschlüssen bei der Interpretation und Auswertung von statistischen Analysen vor, indem sie eine durch den Kontext beziehungsweise durch das Design der Studie bedingte hierarchische Mehrebenenstruktur der Daten berücksichtigen (Eid et al., 2013; Garson, 2013a; Hox, 2010; Luke, 2004). Nach Nezlek et al. (2006) liegt eine solche Mehrebenenstruktur vor, wenn die Daten einer Analyseebene hierarchisch in einer anderen Ebene geschachtelt sind. Dies ist für zwei Arten von Daten charakteristisch: für Daten von Individuen, die bestimmten Gruppen angehören (z. B. Schulklassen), und für Messwiederholungsdaten, bei denen pro Person mehrere Messzeitpunkte existieren (Luke, 2004; Raudenbush & Bryk, 2002). Erstere sind der häufigste Grund für den Einsatz von Mehrebenenanalysen, wobei der Kontext Schule, in welchem Schüler in Klassen und Klassen in Schulen geschachtelt sind (3-Ebenen-Struktur), ein Paradebeispiel hierfür darstellt (Luke, 2004). Letztere zeigen die Datenhierarchie weniger offensichtlich. Messzeitpunkte können als untere Ebene der höheren Ebene, den Individuen, zugeordnet werden (Hox, 2010). Sozialwissenschaftliche Untersuchungen, insbesondere Studien in pädagogischen Kontexten, weisen nicht selten sogar beide hierarchischen Datenmerkmale auf. Oft liegen auf der ersten Ebene Längsschnittdaten mit mehreren Messzeitpunkten pro Individuum vor (zweite Ebene), die wiederum bestimmten Gruppen beziehungsweise Kontextmerkmalen zugeordnet werden können (dritte Ebene) (Ditton, 1998; Raudenbush & Bryk, 2002).

Obwohl gerade in sozialwissenschaftlichen Untersuchungen die Kontextmerkmale beziehungsweise auch sogenannten (*Referenz-*)*Gruppeneffekte*<sup>41</sup> von besonderem Forschungsinteresse sind (z. B. interessiert der Einfluss des Klassenklimas als Gruppenmerkmal auf die schulbezogene Hilfslosigkeit als Individualmerkmal, vgl. die Untersuchung von König (2009) und sich Mehrebenenmodelle als Analyseverfahren anbieten, werden diese nicht selbstverständlich eingesetzt (Ditton, 1998; Raudenbush & Bryk, 2002). Dabei können die Koeffizienten der traditionellen Verfahren wie z. B. Varianz- und Regressionsanalysen nur

---

<sup>41</sup> Ein sehr bekannter und für pädagogisch-psychologische Untersuchungen typischer Kontext- oder auch Referenzgruppeneffekt ist der sogenannte Big-Fish-Little-Pond-Effekt oder auch Fischteicheffekt (Eid et al., 2013; Trautwein & Baeriswyl, 2007). Dieser besagt, dass Schüler gleicher Fähigkeit in Abhängigkeit vom mittleren Fähigkeitsniveau ihrer Klasse unterschiedlich hohe Selbstkonzepte ausbilden. Schüler mit einem bestimmten Fähigkeitsniveau in leistungsstarken Klassen entwickeln ein geringeres Selbstkonzept als Schüler mit gleichem Fähigkeitsniveau in leistungsschwachen Klassen. Das mittlere Leistungsniveau der eigenen Klasse beeinflusst als Kontexteffekt also die Entwicklung des individuellen Selbstkonzepts (Marsh, 2005).

als eine Mischung aus Individual- und Kontexteffekten verstanden werden. Variablen unterschiedlicher Ebenen im Rahmen solcher herkömmlicher Methoden auf einer einzigen Ebene abzubilden, kann erhebliche Trugschlüsse<sup>42</sup> nach sich ziehen (Ditton, 1998; Eid et al., 2013; Garson, 2013a; Hox, 2010; Luke, 2004; Raudenbush & Bryk, 2002). Mehrebenenmodelle bieten einen methodischen Rahmen, um solche Trugschlüsse zu vermeiden: Durch die Möglichkeit, Effekte aller Ebenen sowie sogenannter ‚*cross-level-effects*‘, also Effekte zwischen den Ebenen, simultan zu untersuchen, bilden Mehrebenenanalysen die hierarchischen Strukturen und die daraus resultierenden Abhängigkeiten deutlich besser ab als die traditionellen Verfahren. Darüber hinaus erlauben Mehrebenenanalysen die Modellierung sogenannter *Zufallseffekte* („*random effects*“). Sie berücksichtigen also die zufällige Variation der Regressionskoeffizienten zwischen den Einheiten höherer Ebenen (Ditton, 1998; Garson, 2013a; Nezlek et al., 2006; Raudenbush & Bryk, 2002). Zudem haben Mehrebenenmodelle speziell für die Analyse von Längsschnittdaten den Vorteil, dass auch die Daten von Personen, die nicht zu allen Messzeitpunkten an den Befragungen teilgenommen haben, bei den Schätzungen der Modellparameter berücksichtigt werden können. Dadurch gehen weniger Informationen verloren als bei den herkömmlichen Verfahren, wo die Stichprobe häufig auf die Anzahl an Personen reduziert wird, die an allen Messzeitpunkten teilgenommen haben (Hox, 2010).

In dieser Studie legen der Kontext Schule und das Vorliegen einer *Klumpenstichprobe*<sup>43</sup> eine hierarchische Struktur der zu analysierenden Daten nahe. Schüler sind in Klassen und Klassen in Schulen geschachtelt. Darüber hinaus existieren für jeden Schüler Mehrfachmessungen, sodass sich die Messzeitpunkte als eine weitere hierarchische Ebene modellieren lassen. Um die kontextuelle Hierarchisierung sowie die aus den Mehrfachmessungen entstandenen Abhängigkeiten der Daten zu berücksichtigen und die oben geschilderten Trugschlüsse zu vermeiden, wird hier das Verfahren der Mehrebenenanalyse zur Beantwortung der Fragestellungen herangezogen.

---

<sup>42</sup> Eine Übersicht über typische Trugschlüsse geben Eid et al. (2013) und Hox (2010).

<sup>43</sup> Ein Klumpen bezeichnet eine Gruppe von Personen. Wenn sich eine Stichprobe nicht durch die Zufallsziehung von Einzelpersonen, sondern durch die zufällige Ziehung bestimmter Gruppen oder Klumpen zusammensetzt, wird diese Klumpenstichprobe genannt (Eid et al., 2013). In der vorliegenden Studie bilden die Klassen die Klumpen der Stichprobe, wobei zu beachten ist, dass die Klassen nicht zufällig gezogen wurden, sondern entsprechend der Programmziele nur die siebten Klassen ausgewählt wurden, deren Lehrkräfte zur Projektteilnahme bereit waren.

### 6.5.2 Veränderungsmessung im Rahmen von Mehrebenenanalysen

Veränderung und Wachstum sind z. B. in Form von Lern- und Entwicklungsprozessen feste Bestandteile des alltäglichen Lebens und die Analyse solcher Prozesse ist häufiges Anliegen wissenschaftlicher Untersuchungen. Neben solchen natürlichen Veränderungen gibt es auch bewusst verursachte Veränderungen durch Interventionen, deren Effekte evaluiert werden sollen (Singer & Willett, 2003). Eine solche Evaluation der Wirksamkeit von Interventionen ist Ziel dieser Arbeit. Im Rahmen von Mehrebenenanalysen werden Modelle zur Analyse von Veränderungsprozessen als *Veränderungsmodelle* („*multilevel models of change*“) bezeichnet (Singer & Willett, 2003). Für ein klassisches Prä- und Post-Design werden üblicherweise *kovarianzanalytische Modelle* aufgestellt, in der es keine Ebene der Zeit gibt, sondern der Post-Wert unter Kontrolle des Prä-Wertes auf der Ebene der Schüler vorhergesagt wird. Eine Extraebene für die Messzeitpunkte wird ab drei Messzeitpunkten eingeführt (Hox, 2010). Solche Modelle, die eine eigene Ebene der Zeit mit drei und mehr Messzeitpunkten berücksichtigen, werden auch *Wachstumskurvenmodelle* („*growth curve models*“) genannt. Das innerhalb dieser Kategorie am häufigsten verwendete Modell ist das sogenannte lineare Wachstumskurvenmodell („*linear growth curve model*“). Hierbei wird eine lineare Veränderung über alle Messzeitpunkte angenommen und modelliert (Chou, Yand, Pentz & Hser, 2004; Luhmann & Eid, 2013; Singer & Willett, 2003). Grundsätzlich lassen sich neben der linearen Veränderung auch eine Vielzahl *nicht-linearer Veränderungen* („*non-linear change*“) abbilden (z. B. quadratische oder kubische) (Singer & Willett, 2003). Eine solche Modellierung sogenannter höherer Zeit-Terme kann nach Greenberg und Phillips (2013) jedoch leicht zu einer Überanpassung der Daten führen und sollte deshalb mit großer Vorsicht erfolgen. Höhere Zeit-Terme sollten vor allem dann in das Modell aufgenommen werden, wenn mehr als drei Messungen vorliegen und über die Art der Veränderung konkrete inhaltlich begründete Hypothesen bestehen (Singer & Willett, 2003). Aufgrund des Designs der vorliegenden Studie lässt sich durchaus vermuten, dass die Veränderungen in den Zielkriterien über beide Projektjahre nicht linear verlaufen, sondern sich jahresspezifische Verläufe zeigen. Beispielsweise könnte die Umsetzung der ersten drei Module Klassenklima, Lern- und Leistungsraum und Transparenz im ersten Projektjahr zu einem Anstieg in einem Kriterium führen und die Umsetzung der letzten drei Module des FOSIS2-Programms Selbstbestimmung, Kooperatives Lernen und Selbstregulation einen Abfall oder eine Stagnation desselben Kriteriums nach sich ziehen. Wegen der geringen Anzahl an Messzeitpunkten (drei) liegen jedoch keine konkreten Hypothesen über eine besondere Form der Veränderungen für die einzelnen Jahre vor, sodass eine Modellierung



von spezifischen höheren Zeit-Termen, z. B. bestimmten kurvenförmigen Zeitverläufen, nicht möglich ist. Auch wenn keine solchen Vorannahmen über die Form des Verlaufs getroffen und modelliert werden können, gibt es aber durchaus Modellbildungen, die eine Prüfung der differenziellen jahresspezifischen Effekte in einem Modell ermöglichen, ohne spezifische höhere Zeit-Terme zu modellieren. Solche Modelle werden auch *diskontinuierliche Veränderungsmodelle* genannt (*„discontinuous change models“*) (Luhmann & Eid, 2013; Singer & Willett, 2003). Muthén (2001) spricht auch von *Phasen- oder sequentiellen Veränderungsmodellen* (*„piecewise or sequential change models“*).

In der vorliegenden Untersuchung soll zur Erfassung der Veränderung in den motivational relevanten Zielkriterien zunächst ein Drei-Ebenen-Modell mit der Zeit als eigene Ebene aufgestellt und sowohl die kontextuelle Hierarchisierung, also die Schachtelung von Schülern in Klassen, als auch die aus den Mehrfachmessungen entstandenen Abhängigkeiten berücksichtigt werden. Dabei bilden die Messzeitpunkte die erste, die Schüler die zweite und die Klassen die dritte Ebene. Aufgrund der geringen Anzahl untersuchter Schulen ( $N = 6$ ) kann die höchste Kontext-Ebene „Schule“ nicht in die Mehrebenenmodellierung eingeschlossen werden (siehe auch Anhang C 1 „Voraussetzungen für die Anwendung von Mehrebenenanalysen“). Da die Messung von Veränderungen über alle Messzeitpunkte hinweg mittels linearer Wachstumskurvenmodelle die übliche Form der Modellbildung, also *„the state of the art“* darstellt (Greenberg & Phillips, 2013), werden in dieser Arbeit zur Analyse der Frage nach der Wirksamkeit des Programms zunächst solche linearen Wachstumskurvenmodelle berechnet. Wenn jedoch das Design der Studie wie im vorliegenden Fall phasen-spezifische Effekte nahelegt, sollten diese nach Luhmann und Eid (2013) auch unbedingt untersucht werden. Deshalb werden im Anschluss an die Analyse der linearen Verläufe über beide Projektjahre hinweg die jahresspezifischen Effekte mittels der oben benannten diskontinuierlichen Veränderungsmodelle geprüft.

Die Frage nach der Wirksamkeit in Abhängigkeit von der Implementation soll beispielhaft für das erste Projektjahr untersucht werden (siehe Abschnitt 6.5.4.) Für die entsprechenden Analysen werden folglich nur der erste und zweite Messzeitpunkt herangezogen, die Ebene der Messzeitpunkte entfällt und es wird ein Zwei-Ebenen-Modell mit den Schülern auf der ersten und den Klassen auf der zweiten Ebene, modelliert. Für ein solches klassisches Prä- und Post-Design werden üblicherweise kovarianzanalytische Modelle herangezogen, bei denen der Post-Wert unter Kontrolle des Prä-Wertes auf der Ebene der Schüler vorhergesagt wird. Solche kovarianzanalytischen Modelle werden auch in dieser Arbeit berechnet.

Da die Kontrollschüler nicht zum Interventionsgeschehen befragt werden können, beschränken sich die Analysen für die Frage nach der Wirksamkeit in Abhängigkeit von der Implementation auf die Interventionsgruppe.

Nachfolgend werden die in dieser Studie durchgeführten Modelle erläutert. Die Modellbildung erfolgt dabei durch eine sukzessive Aufnahme der Prädiktoren.

### 6.5.3 Mehrebenenmodelle zur Evaluation der Wirksamkeit

Für die Durchführung von Mehrebenenmodellen mit der Zeit als Ebene muss der SPSS-Datensatz zunächst anders strukturiert werden. Die Messzeitpunkte einer Person sind hier nicht mehr horizontal, sondern vertikal geordnet. Dadurch wird die Zuordnung von Messzeitpunkten als unterste Ebene zur nächsthöheren Ebene, den Schülern, deutlich (Abbildung 9).

Singer & Willet (2003) empfehlen bei der Modellierung von Längsschnittdaten mit den Messzeitpunkten als unterste Ebene, zunächst mit dem schlichtesten Modell, dem Nullmodell ohne jegliche Prädiktoren, zu beginnen. Dieses Modell beschreibt noch keine Veränderung, sondern nur die Variation der abhängigen Variablen über alle Messzeitpunkte und Gruppen hinweg und dient der Ermittlung substantieller Varianz und der sogenannten Intraklassen-Korrelation (eine Erläuterung der substantiellen Varianz und der Intraklassen-Korrelationen findet sich im Anhang unter C1).

#### Nullmodell

##### *Level 1: Ebene der Messzeitpunkte*

$$Y_{tij} = \pi_{0ij} + e_{tij}$$

$Y_{tij}$  abhängige Variable, Wert eines Schülers i in Klasse j zum Zeitpunkt t

$\pi_{0ij}$  Mittelwert eines Schülers i in Klasse j über alle Messzeitpunkte

$e_{tij}$  Abweichung des beobachteten Werts eines Schülers i in Klasse j vom vorhergesagten Wert zum Messzeitpunkt t

##### *Level 2: Schülerebene*

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$\beta_{00j}$  Mittelwert der Klasse j über alle Messzeitpunkte

$r_{0ij}$  Abweichung des Schülers i der Klasse j vom Mittelwert der Klasse j über alle Messzeitpunkte

*Level 3: Klassenebene*

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$\gamma_{000}$  Mittelwert aller Klassen über alle Messzeitpunkte

$u_{00j}$  Abweichung der Klasse j vom Mittelwert aller Klassen über alle Messzeitpunkte

Im Anschluss an das Nullmodell werden zur Evaluation der Wirksamkeit des FOSIS2-Programms zwei verschiedene Modelltypen gebildet, deren Submodelle jeweils ineinander geschachtelt sind: lineare Wachstumskurvenmodelle zur Evaluation der Wirksamkeit über beide Projektjahre hinweg und diskontinuierliche Veränderungsmodelle zur Prüfung jahresspezifischer Effekte.

***Lineare Wachstumskurvenmodelle***

Um den linearen Verlauf über beide Projektjahre zu modellieren, ist die Bildung einer kontinuierlichen Zeitvariablen ‚Zeit‘ nötig. Diese wird mit „Null“ für den ersten, „Eins“ für den zweiten und „Zwei“ für den dritten Messzeitpunkt kodiert. Abbildung 9 zeigt einen Ausschnitt aus einem umstrukturierten SPSS-Datensatz mit der Variablen Zeit und Skalenwerten für die Variablen wahrgenommene Bedrohung und Prüfungsängstlichkeit, jeweils abgekürzt mit ‚thre‘ für ‚threat‘ und ‚taiw‘ für ‚test anxiety‘.

id	thre	taiw	Zeit
9	3,00	3,80	0
9	1,33	2,80	1
9	1,33	2,60	2
10	2,33	2,20	0
10	2,67	2,60	1
10	1,33	3,40	2
11	3,00	3,20	0
11	1,33	3,40	1
11	2,00	3,00	2

**Abbildung 9:** Screenshot des umstrukturierten SPSS-Datensatzes mit einer kontinuierlichen Zeitvariablen

Der Schüler mit der ID 9 weist in der Prüfungsängstlichkeit zum ersten Messzeitpunkt den Wert 3.8, zum zweiten Messzeitpunkt den Wert 2.8 und zum dritten Messzeitpunkt den Wert 2.6 auf. Durch die Abbildung 9 lässt sich auch die Schachtelung oder Gruppierung von Messzeitpunkten in einem Schüler nachvollziehen: Für jeden Schüler liegt eine Gruppe von drei Messzeitpunkten vor, die untereinander angeordnet sind.

Nach dem Nullmodell wird ein Modell aufgestellt, in dem zunächst die Zeitvariable auf der ersten Ebene aufgenommen wird. Dieses Modell bildet die Veränderung über die Zeit ab,

also über beide Projektjahre hinweg, und wird deshalb ‚Veränderungsmodell (Zeit)‘ genannt. Das Modell gibt noch keine Auskunft darüber, inwieweit sich Interventions- und Kontrollgruppe unterscheiden.

### Veränderungsmodell (Zeit)

#### Level 1: Ebene der Messzeitpunkte

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Zeit}) + e_{tij}$$

$Y_{tij}$  abhängige Variable, Wert eines Schülers i in Klasse j zum Zeitpunkt t

$\pi_{0ij}$  Wert eines Schülers i in Klasse j zum ersten Messzeitpunkt

$\pi_{1ij}$  lineare Veränderung eines Schülers i in Klasse j über die Zeit

$e_{tij}$  Abweichung des beobachteten Werts eines Schülers i in Klasse j vom vorhergesagten Wert zum Messzeitpunkt t

#### Level 2: Schülerebene

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$\beta_{00j}$  Mittelwert der Klasse j zum ersten Messzeitpunkt

$\beta_{10j}$  lineare mittlere Veränderung der Klasse j über die Zeit

$r_{0ij}$  Abweichung des Schülers i der Klasse j vom Mittelwert der Klasse j zum ersten Messzeitpunkt

$r_{1ij}$  Abweichung der Veränderung des Schülers i der Klasse j von der mittleren Veränderung der Klasse j über die Zeit

#### Level 3: Klassenebene

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$\gamma_{000}$  Mittelwert aller Klassen zum ersten Messzeitpunkt

$\gamma_{100}$  lineare mittlere Veränderung aller Klassen über die Zeit

$u_{00j}$  Abweichung der Klasse j vom Mittelwert aller Klassen zum ersten Messzeitpunkt

$u_{10j}$  Abweichung der Veränderung der Klasse j von der mittleren Veränderung aller Klassen über die Zeit

In einem nächsten Modell wird neben der Zeitvariablen die Zugehörigkeit zur Interventions- und Kontrollgruppe modelliert. Da die Gruppenzugehörigkeit ein Klassenmerkmal darstellt, wird diese – hier gekennzeichnet mit ‚Gruppe‘ – als Level-3-Merkmal aufgenommen. Die Variable ist mit „Null“ für die Kontrollgruppe und „Eins“ für die Interventionsgruppe kodiert. Mit diesem Modell wird die Veränderung der Interventionsgruppe über

die Zeit im Vergleich zur Kontrollgruppe untersucht. Es handelt sich folglich um das Zielmodell, das zur Beantwortung der ersten Fragestellung nach der Wirksamkeit des FOSIS2-Programms über die Zeit, also über beide Projektjahre hinweg, herangezogen wird. Es wird als ‚Evaluationsmodell (Zeit & Gruppe)‘ bezeichnet.

### **Evaluationsmodell (Zeit & Gruppe)**

#### *Level 1: Ebene der Messzeitpunkte*

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Zeit}) + e_{tij}$$

Die Bedeutung der Koeffizienten entspricht der im Veränderungsmodell (T1-T3).

#### *Level 2: Schülerebene*

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Die Bedeutung der Koeffizienten entspricht der im Veränderungsmodell (T1-T3).

#### *Level 3: Klassenebene*

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Gruppe}) + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{Gruppe}) + u_{10j}$$

$\gamma_{000}$  Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt

$\gamma_{001}$  Unterschied zwischen dem Mittelwert aller Interventionsklassen und dem Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt

$\gamma_{100}$  lineare mittlere Veränderung aller Kontrollklassen über die Zeit

$\gamma_{101}$  Unterschied zwischen der linearen mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen über die Zeit

$u_{00j}$  Abweichung der Klasse j zum ersten Messzeitpunkt vom jeweiligen Gruppenmittelwert (Interventions- und Kontrollgruppe)

$u_{10j}$  Abweichung der Veränderung der Klasse j von der mittleren Veränderung der jeweiligen Gruppe (Interventions- und Kontrollgruppe) über die Zeit

Die Bedeutung und Interpretation der Koeffizienten des Zielmodells ist am besten nachzuvollziehen, wenn die Bildung der sogenannten Gesamtgleichung erfolgt. Dazu werden die unteren beiden Ebenen-Gleichungen in die erste Ebenen-Gleichung eingesetzt. Zur Vereinfachung der Darstellung und zur Erhöhung der Verständlichkeit werden die Zufallseffekte weggelassen.

### **Gesamtgleichung Evaluationsmodell (Zeit & Gruppe) ohne Zufallseffekte**

$$Y_{tij} = \gamma_{000} + \gamma_{001}(\text{Gruppe}) + \gamma_{100}(\text{Zeit}) + \gamma_{101}(\text{Zeit} \times \text{Gruppe})$$

Durch das Einsetzen der möglichen Kodierungen der Zeitvariablen „Null“ (T1), „Eins“ (T2) und „Zwei“ (T3) sowie der Variablen ‚Gruppe‘ mit „Null“ (Kontrollgruppe) und „Eins“ (Interventionsgruppe) in diese um die Zufallseffekte gekürzte Gesamtgleichung lässt sich die Bedeutung der festen Effekte des Zielmodells ermitteln. Der erste Parameter  $\gamma_{000}$  bildet den mittleren Wert aller Kontrollklassen zum ersten Messzeitpunkt, also zur Präbefragung ab. Den Unterschied in den Ausgangswerten, also die Differenz zwischen dem Mittelwert aller Interventionsklassen und dem Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt erfasst der zweite Koeffizient  $\gamma_{001}$ . An dritter Stelle beschreibt  $\gamma_{100}$  die durchschnittliche Veränderung aller Kontrollklassen über die Zeit. Der Interaktionseffekt  $\gamma_{101}$  ist der für die Beantwortung der Forschungsfrage wichtigste Effekt. Dieser verdeutlicht die Abweichung der mittleren Veränderung aller Interventionsklassen über die Zeit von der (mittleren) Veränderung aller Kontrollklassen und zeigt somit den Unterschied in der Veränderung über die Zeit zwischen den Interventions- und Kontrollklassen auf.

In HLM 6.08 kann definiert werden, welche Zufallseffekte modelliert und welche fixiert werden sollen. Neben inhaltlichen Überlegungen werden die im Anhang C2 dargestellten Kriterien zur Beurteilung der Modellgüte (Pseudo- $R^2$ , Chi-Quadrat-Test für den Devianzvergleich und „AIC – Akaike Information Criterion“) herangezogen, um über die Zufallseffekte zu entscheiden: Da auf dem zweiten Schüler-Level keine Prädiktoren aufgenommen werden sollen, interessiert die Varianzaufklärung auf dieser Ebene weniger als auf der dritten, wo zur Beantwortung der Forschungsfrage die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ erfolgt. Die Zufallseffekte der zweiten Ebene könnten folglich fixiert und die Varianzkomponenten auf der dritten Ebene sollten modelliert werden. Im Hinblick auf die Frage nach der Wirksamkeit der Intervention in Abhängigkeit von der individuell wahrgenommenen Implementation seitens der Schüler ist jedoch die Varianz zwischen den Schülern einer Klasse durchaus von Interesse. Denn die Implementation wird später in den kovarianzanalytischen Modellen als Individualmerkmal auf der Schülerebene aufgenommen. Um explorativ erste Eindrücke über die Varianz zwischen den Schülern einer Klasse zu erhalten, werden die Zufallseffekte auf der Schülerebene bereits hier in den Wachstumskurvenmodellen modelliert und dürfen variieren. Neben diesen inhaltlichen Überlegungen geben auch die Kriterien zur Beurteilung der Modellgüte keinen Anlass, die Zufallseffekte auf der Schülerebene zu fixieren (siehe auch Kapitel 7). Entsprechend werden

die Zufallseffekte aller Ebenen modelliert, jedoch die Varianzaufklärung mittels des Pseudo- $R^2$  nur für die dritte Ebene, auf der Prädiktoren aufgenommen werden, berechnet.

### ***Diskontinuierliche Veränderungsmodelle***

Zur Überprüfung phasen-spezifischer oder auch sequentieller Jahres-Effekte sollen die oben beschriebenen diskontinuierlichen oder auch sogenannten sequentiellen beziehungsweise Phasen-Veränderungsmodelle eingesetzt werden (Luhmann & Eid, 2013; Muthén, 2001; Singer & Willett, 2003). Anders als zuvor bei der Modellierung der linearen Wachstumskurvenmodelle wird die Veränderung der Interventionsgruppe im Vergleich zur Kontrollgruppe nun für beide Projektjahre einzeln untersucht. Hierbei wird die Zeit nicht mehr linear durch eine kontinuierliche Zeitvariable abgebildet, sondern durch zwei diskontinuierliche, dummy-kodierte Zeit-Variablen, die jeweils die Veränderung über die einzelnen Projektjahre beschreiben:

Um die Veränderung vom ersten zum zweiten Messzeitpunkt zu messen, wird die Variable *„J1“ (Jahr 1)* gebildet. Hier wird der erste Messzeitpunkt mit „Null“, der zweite mit „Eins“ und der dritte ebenfalls mit „Eins“ kodiert. Für die Veränderung vom zweiten zum dritten Projektjahr wird die Variable *„J2“ (Jahr 2)* aufgenommen, wobei hier der erste Messzeitpunkt mit „Null“, der zweite ebenfalls mit „Null“ und der dritte mit „Eins“ kodiert wird (siehe Abbildung 10). Eine solche Kodierung wird auch *„neighbouring Change“* (Luhmann & Eid, 2013) genannt. Denn durch diese Art der Dummy-Kodierung bilden die beiden Zeitvariablen *„J1“* und *„J2“* die benachbarten Jahresveränderungen ab. Es wären auch andere Formen von Dummy-Kodierungen möglich, durch die beispielsweise die zweite Zeitvariable nicht die benachbarte Veränderung zum ersten Projektjahr erfassen würde, sondern die Veränderung vom ersten zum dritten Messzeitpunkt. Diese Art der Kodierung wird auch *„baseline Change“* genannt (Luhmann & Eid, 2013). Ergänzend zu den linearen Wachstumsmodellen sollen in dieser Arbeit mithilfe der diskontinuierlichen Veränderungsmodelle die benachbarten jahresspezifischen Veränderungen untersucht werden. Entsprechend wird die geschilderte Kodierung des *„neighbouring Change“* verwendet. Abbildung 10 verdeutlicht in einem Ausschnitt aus einem SPSS-Datensatz die Kodierung der Zeitvariablen *„J1“* und *„J2“*.

id	thre	taiw	J1	J2
9	3,00	3,80	0	0
9	1,33	2,80	1	0
9	1,33	2,60	1	1
10	2,33	2,20	0	0
10	2,67	2,60	1	0
10	1,33	3,40	1	1
11	3,00	3,20	0	0
11	1,33	3,40	1	0
11	2,00	3,00	1	1

**Abbildung 10: Screenshot des umstrukturierten SPSS-Datensatzes mit zwei Dummy-Zeitvariablen**

Wie bei den linearen Modellen wird auch bei diskontinuierlichen Veränderungsanalysen mit dem bereits beschriebenen Nullmodell begonnen, um die Intraklassen-Korrelation zu ermitteln und zu überprüfen, ob substantielle Level-2 beziehungsweise Level-3-Varianz vorhanden ist.

Im nächsten Modell wird wiederum die Veränderung über die Zeit, diesmal jedoch jeweils getrennt für das erste und das zweite Projektjahr modelliert. Anstelle der kontinuierlichen Zeitvariablen werden die bereits erläuterten Dummy-Variablen aufgenommen. Dieses Modell wird hier als ‚Veränderungsmodell (J1 & J2)‘ bezeichnet.

### Veränderungsmodell (J1 & J2)

#### Level 1: Ebene der Messzeitpunkte

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} (J1) + \pi_{2ij} (J2) + e_{tij}$$

$Y_{tij}$  abhängige Variable, Wert eines Schülers i in Klasse j zum Zeitpunkt t

$\pi_{0ij}$  Wert eines Schülers i in Klasse j zum ersten Messzeitpunkt

$\pi_{1ij}$  Veränderung eines Schülers i in Klasse j vom ersten zum zweiten Messzeitpunkt

$\pi_{2ij}$  Veränderung eines Schülers i in Klasse j vom zweiten zum dritten Messzeitpunkt

$e_{tij}$  Abweichung des beobachteten Werts eines Schülers i in Klasse j vom vorhergesagten Wert zum Messzeitpunkt t (messzeitpunktspezifischer Fehlerterm)

#### Level 2: Schülerebene

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + (r_{2ij})^{44}$$

<sup>44</sup> Die Varianzkomponente  $r_{2ij}$  ist eingeklammert, weil sie in diesem und auch im nächsten Modell fixiert wird. Eine ausführliche Begründung dieses Vorgehens erfolgt im Anschluss an die Beschreibung des Evaluationsmodells (J1, J2 & Gruppe).



$\beta_{00j}$	Mittelwert der Klasse j zum ersten Messzeitpunkt
$\beta_{10j}$	mittlere Veränderung der Klasse j vom ersten zum zweiten Messzeitpunkt
$\beta_{20j}$	mittlere Veränderung der Klasse j vom zweiten zum dritten Messzeitpunkt
$r_{0ij}$	Abweichung des Schülers i der Klasse j vom Mittelwert der Klasse j zum ersten Messzeitpunkt
$r_{1ij}$	Abweichung der Veränderung des Schülers i der Klasse j von der mittleren Veränderung der Klasse j vom ersten zum zweiten Messzeitpunkt
$r_{2ij}$	Abweichung der Veränderung des Schülers i in Klasse j von der mittleren Veränderung der Klasse j vom zweiten zum dritten Messzeitpunkt

*Level 3: Klassenebene*

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{20j} = \gamma_{200} + u_{20j}$$

$\gamma_{000}$	Mittelwert aller Klassen zum ersten Messzeitpunkt
$\gamma_{100}$	mittlere Veränderung aller Klassen vom ersten zum zweiten Messzeitpunkt
$\gamma_{200}$	mittlere Veränderung aller Klassen vom zweiten zum dritten Messzeitpunkt
$u_{00j}$	Abweichung der Klasse j vom Mittelwert aller Klassen zum ersten Messzeitpunkt
$u_{10j}$	Abweichung der Veränderung der Klasse j von der mittleren Veränderung aller Klassen vom ersten zum zweiten Messzeitpunkt
$u_{20j}$	Abweichung der Veränderung der Klasse j von der mittleren Veränderung aller Klassen vom zweiten zum dritten Messzeitpunkt

Wie auch das lineare Veränderungsmodell gibt dieses diskontinuierliche Veränderungsmodell noch keine Auskunft über die Unterschiede zwischen Interventions- und Kontrollklassen, sondern beschreibt zunächst nur die jahresspezifische Veränderung über alle Klassen.

Erst im dritten Modell, dem sogenannten ‚Evaluationsmodell (J1, J2 & Gruppe)‘, soll die Veränderung der Interventionsgruppe im Vergleich zur Kontrollgruppe jeweils für die beiden Projektjahre getrennt untersucht werden. Hierzu wird wiederum die Variable ‚Gruppe‘, kodiert mit „Null“ für die Kontroll- und „Eins“ für die Interventionsklassen, auf der Klassenebene aufgenommen. Dieses ‚diskontinuierliche Evaluationsmodell (J1, J2 & Gruppe)‘ ist neben dem ‚linearen Evaluationsmodell (Zeit & Gruppe)‘ ein weiteres Zielmodell, das zur Beantwortung der ersten Fragestellung, speziell zur Überprüfung jahresspezifischer Gruppeneffekte, herangezogen wird.

**Evaluationsmodell (J1, J2 & Gruppe)***Level 1: Ebene der Messzeitpunkte*

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} (J1) + \pi_{2ij} (J2) + e_{tij}$$

Die Bedeutung der Koeffizienten entspricht der im Veränderungsmodell (J1 & J2).

*Level 2: Schülerebene*

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + (r_{2ij})$$

Die Bedeutung der Koeffizienten entspricht der im Veränderungsmodell (J1 & J2).

*Level 3: Klassenebene*

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Gruppe}) + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(\text{Gruppe}) + u_{10j}$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}(\text{Gruppe}) + u_{20j}$$

$\gamma_{000}$  Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt

$\gamma_{100}$  mittlere Veränderung aller Kontrollklassen vom ersten zum zweiten Messzeitpunkt

$\gamma_{200}$  mittlere Veränderung aller Kontrollklassen vom zweiten zum dritten Messzeitpunkt

$\gamma_{001}$  Unterschied zwischen dem Mittelwert aller Interventionsklassen und dem Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt

$\gamma_{101}$  Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen vom ersten zum zweiten Messzeitpunkt

$\gamma_{201}$  Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen vom zweiten zum dritten Messzeitpunkt

$u_{00j}$  Abweichung der Klasse j zum ersten Messzeitpunkt vom jeweiligen Gruppenmittelwert (Interventions- und Kontrollgruppe)

$u_{10j}$  Abweichung der Veränderung der Klasse j von der mittleren Veränderung der jeweiligen Gruppe (Interventions- und Kontrollgruppe) vom ersten zum zweiten Messzeitpunkt

$u_{20j}$  Abweichung der Veränderung der Klasse j von der mittleren Veränderung der jeweiligen Gruppe (Interventions- und Kontrollgruppe) vom zweiten zum dritten Messzeitpunkt

Die Koeffizienten dieses Modells lassen sich wiederum besser nachvollziehen, wenn die Gesamtgleichung betrachtet wird. Zur Vereinfachung der Darstellung und der besseren Verständlichkeit werden die Zufallseffekte erneut weggelassen.

**Gesamtgleichung Evaluationsmodell (J1, J2 und Gruppe) ohne Zufallseffekte**

$$Y_{tij} = \gamma_{000} + \gamma_{001}(\text{Gruppe}) + \gamma_{100}(J1) + \gamma_{101}(J1 \times \text{Gruppe}) + \gamma_{200}(J2) + \gamma_{201}(J2 \times \text{Gruppe})$$

Durch das Einsetzen der möglichen Kodierungen der Gruppenvariablen (0 = Kontrollgruppe, 1 = Interventionsgruppe) sowie der Zeitvariablen ‚J1‘ (0 = T1, 1 = T2, 1 = T3) und ‚J2‘ (0 = T1, 0 = T2, 1 = T3) in diese um die Zufallseffekte gekürzte Gesamtgleichung lässt sich die Bedeutung der festen Effekte des Zielmodells ermitteln. Auch hier beschreibt  $\gamma_{000}$  den mittleren Wert aller Kontrollklassen zum ersten Messzeitpunkt und  $\gamma_{001}$  wiederum die Abweichung der Interventionsklassen von den Kontrollklassen zum ersten Messzeitpunkt. Der Koeffizient  $\gamma_{001}$  erfasst folglich den Unterschied in den Ausgangswerten zwischen der Interventions- und der Kontrollgruppe. Über die durchschnittliche Veränderung aller Kontrollklassen vom ersten zum zweiten Messzeitpunkt gibt  $\gamma_{100}$  Auskunft. Der für die Fragestellung interessierende Interaktionseffekt ‚J1 x Gruppe‘ für das erste Projektjahr wird durch  $\gamma_{101}$  modelliert. Dieser Koeffizient bildet die Abweichung der Interventions- von der Kontrollgruppe in der mittleren Veränderung über das erste Projektjahr ab. Die beiden Parameter  $\gamma_{200}$  und  $\gamma_{201}$  sind analog zu  $\gamma_{100}$  und  $\gamma_{101}$  für das zweite Projektjahr zu interpretieren. Durch  $\gamma_{200}$  wird die durchschnittliche Veränderung aller Kontrollklassen vom zweiten zum dritten Messzeitpunkt erfasst und  $\gamma_{201}$  beschreibt den zweiten für die Fragestellung interessierenden Interaktionseffekt ‚J2 x Gruppe‘: die Abweichung der Interventions- von der Kontrollgruppe in der mittleren Veränderung über das zweite Projektjahr.

Hinsichtlich der Frage nach der Fixierung oder Modellierung der Zufallseffekte verschiedener Ebenen gelten für die diskontinuierlichen Modelle dieselben Überlegungen wie für die linearen Analysen: Bis auf die eingeklammerte Schüler-Varianzkomponente  $r_{2ij}$ , die sich auf die Veränderung im zweiten Projektjahr bezieht, werden alle Varianzkomponenten modelliert. Die Varianzkomponente  $r_{2ij}$  wird sowohl im Veränderungs- als auch im Evaluationsmodell fixiert, da für die Schätzung der Modellparameter sonst nicht genügend Freiheitsgrade vorhanden wären, um eine Mehrebenenanalyse mit HLM 6.08 durchzuführen.<sup>45</sup> Zudem werden alle Varianzkomponenten der dritten Klassenebene modelliert, da diese durch den Prädiktor ‚Gruppe‘ aufgeklärt und die Varianzaufklärung mittels Pseudo- $R^2$  ermittelt werden soll. Letzteres ist nur durch eine Modellierung der Zufallseffekte dieser Ebene möglich. Die Kriterien zur Beurteilung der Modellgüte bestätigen das hier be-

---

<sup>45</sup> Ohne eine Fixierung von  $r_{2ij}$  erscheint bei HLM 6.08 folgende Fehlermeldung: „There are no degrees of freedom to estimate sigma\_squared. Set sigma\_squared to a constant, or delete one or more of the random effects from the model“. In den HLM-Hilfen Raudenbush, Bryk und Congdon (2009) wird für den Umgang mit dieser Fehlermeldung empfohlen, möglichst solche Varianzkomponenten zu fixieren beziehungsweise aus dem Modell zu entfernen, die sehr gering ausfallen. Da sich über alle hier untersuchten Kriterien hinweg die Varianz  $r_{2ij}$  als sehr gering erweist, wird diese Komponente in allen Modellen fixiert.

schriebene Vorgehen, denn die Modelle mit allen Zufallseffekten auf der dritten Klassenebene und den Zufallseffekten für den Achsenabschnitt und die Steigung des ersten Projektjahres haben die besten Fit-Indizes (siehe auch Kapitel 7).

#### **6.5.4 Mehrebenenmodelle zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation**

Neben der linearen und jahresspezifischen Evaluation der Wirksamkeit des FOSIS2-Programms soll auch der Effekt der wahrgenommenen Implementation seitens der Schüler auf die Veränderung in den Kriterien überprüft werden. Dabei wird der Einfluss der individuell wahrgenommenen Implementation als ein Personenmerkmal, also Level-1-Prädiktor, modelliert. Da anzunehmen ist, dass die Schüler einer Klasse ein ähnliches Maß an Implementation erleben, wäre eine andere inhaltlich sinnvolle Form der Modellierung die Aggregation der einzelnen Schülerwerte zu einem Klassenwert und die Aufnahme dieser aggregierten Implementationsvariablen auf der Klassenebene gewesen. In anderen im FOSIS2-Projekt angesiedelten Untersuchungen werden solche individuellen und aggregierten Schüler- sowie zusätzlich auch Lehrerimplementationsangaben hinsichtlich ihrer Vorhersagefähigkeit validiert. Die individuell wahrgenommene Implementation der Schüler hat sich hierbei als validester Indikator für die Implementation erwiesen und wird deshalb auch in der vorliegenden Studie verwendet.

Für jedes der sechs Module des FOSIS2-Programms stehen eigene Implementationsindikatoren zur Verfügung, für die in jeweils unterschiedlichen Kombinationen zahlreiche Mehrebenenmodelle zur Untersuchung des Einflusses der Implementation denkbar wären. Die sukzessive Aufnahme aller sechs Implementationsindikatoren führt zu sehr komplexen Modellen, für die eine hohe Anzahl an Parametern geschätzt werden muss. Solche komplexen Modelle verursachen bei der hier verwendeten kleinen Stichprobengröße Konvergenz- und Iterationsprobleme. Um diese zu vermeiden und die Möglichkeiten der Modellbildung zu reduzieren, wird die Frage nach der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation hier beispielhaft für das erste Projektjahr und die ersten drei Module untersucht. Die Entscheidung für das erste Projektjahr lässt sich mithilfe der deskriptiven Betrachtung der Implementation (siehe Abschnitt 7.2.1) begründen: Die Beteiligung der Lehrkräfte an den Workshops, die Inanspruchnahme von Beratungen und die von den Schülern wahrgenommene modulspezifische Implementation legen nahe, dass das tatsächliche Umsetzungsverhalten der Lehrkräfte im ersten Projektjahr höher ausfällt als im zweiten. Ausgehend von der Annahme, dass die Quantität der Umsetzung einen Ein-

fluss auf die Programmwirksamkeit hat (Durlak & DuPre, 2008), lässt sich somit vermuten, dass sich insbesondere für die Maßnahmen des ersten Projektjahres Zusammenhänge zwischen der Implementation und den Zielkriterien zeigen.

Da sich die Modelle zur Untersuchung der Implementationswahrnehmung folglich ausschließlich auf das erste Projektjahr beziehen, werden nur der erste und zweite Messzeitpunkt in die Modellbildung einbezogen. Dadurch entfällt die Ebene der Messzeitpunkte und es wird ein 2-Ebenen-Modell mit den Schülern auf der ersten und den Klassen auf der zweiten Ebene modelliert. Für ein solches klassisches Prä-Post-Design werden üblicherweise kovarianzanalytische Modelle herangezogen, bei denen der Post-Wert unter Kontrolle des Prä-Wertes auf der Ebene der Schüler vorhergesagt wird.

Wie bei dem linearen und diskontinuierlichen Vorgehen wird auch zur Analyse der Wirksamkeit in Abhängigkeit von der Implementation für das erste Projektjahr zunächst ein Nullmodell aufgestellt, um zu überprüfen, ob die Intraklassen-Korrelation ein mehr Ebenenanalytisches Vorgehen nahelegt. Dieses Nullmodell unterscheidet sich vom bereits beschriebenen Nullmodell in der Anzahl der Ebenen (nur zwei) und der herangezogenen Stichprobe (ausschließlich Interventionsschüler).

#### **Nullmodell**

##### *Level 1: Schülerebene*

$$Y_{ij} = \beta_{0j} + r_{ij}$$

$Y_{ij}$  Wert eines Schülers  $i$  in Klasse  $j$  zum zweiten Messzeitpunkt (abhängige Variable)

$\beta_{0j}$  Mittelwert der Klasse  $j$  zum zweiten Messzeitpunkt (klassenspezifischer Achsenabschnitt)

$r_{ij}$  Abweichung des Schülers  $i$  von seinem klassenspezifischen Achsenabschnitt

##### *Level 2: Klassenebene*

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$\gamma_{00}$  Mittelwert aller Klassen zum zweiten Messzeitpunkt (mittlerer Achsenabschnitt)

$u_{0j}$  Abweichung der Klasse  $j$  vom mittleren Achsenabschnitt

#### **Kovarianzanalytische Veränderungsmodelle**

Im Anschluss an das Nullmodell werden in einem nächsten Modell zunächst die *Ausgangswerte* des jeweiligen Zielkriteriums als Level-1-Prädiktor, hier gekennzeichnet mit  $TI'$ , auf der ersten Ebene aufgenommen. Die Variable wird um den Gruppenmittelwert zentriert. Diese Art der Zentrierung eignet sich besonders gut für die Analyse von Fragestellungen, bei denen die Variation zwischen den Clustern auf der höheren Ebene (hier die

Variation zwischen den Klassen) nicht interessiert. Beim ‚Group-Mean-Centering‘ wird die Variation zwischen den Gruppen nivelliert und somit aus dem Modell herausgerechnet (Eid et al., 2013). Da der Einfluss der von den Schülern wahrgenommenen Implementation als ein Level-1-Merkmal modelliert wird und keine Level-2-Prädiktoren in das Modell aufgenommen werden, ist die Zentrierung um die Gruppenmittelwerte ein angemessenes Vorgehen. Das Modell mit der Variablen ‚T1‘ als Level-1-Prädiktor wird hier als ‚Autoregressives Modell (T1)‘ bezeichnet. Nach Eid et al. (2013) sind Modelle autoregressiv, wenn „in ihnen die Variation eines Merkmals auf die Variation desselben Merkmals zu einem früheren Messzeitpunkt zurückgeführt wird“ (S. 931).

### **Autoregressives Modell (T1)**

#### *Level 1: Schülerebene*

$$Y_{ij} = \beta_{0j} + \beta_{1j}T1 + r_{ij}$$

$Y_{ij}$  Wert eines Schülers i in Klasse j zum zweiten Messzeitpunkt (abhängige Variable)

$\beta_{0j}$  erwarteter Wert für die Schüler der Klasse j zum zweiten Messzeitpunkt, deren Ausprägungen in der Variablen ‚T1‘ dem Klassenmittelwert entsprechen (klassenspezifischer Achsenabschnitt)

$\beta_{1j}$  Einfluss (Kontrolle) der Variablen ‚T1‘ auf die Werte zum zweiten Messzeitpunkt für Schüler der Klasse j (klassenspezifischer Steigungskoeffizient ‚T1‘)

$r_{ij}$  Abweichung des beobachteten Werts eines Schülers i in Klasse j von seinem vorhergesagten Wert zum zweiten Messzeitpunkt

#### *Level 2: Klassenebene*

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + (u_{1j})^{46}$$

$\gamma_{00}$  erwarteter Wert für die Schüler aller Klassen zum zweiten Messzeitpunkt, deren Ausprägungen in der Variablen ‚T1‘ ihren jeweiligen Klassenmittelwerten entsprechen (mittlerer Achsenabschnitt)

$\gamma_{10}$  mittlerer Einfluss (Kontrolle) der Variablen ‚T1‘ auf die Werte zum zweiten Messzeitpunkt für Schüler aller Klassen (mittlerer Steigungskoeffizient ‚T1‘)

$u_{0j}$  Abweichung der Klasse j vom mittleren Achsenabschnitt

$u_{1j}$  Abweichung der Klasse j vom mittleren Steigungskoeffizienten ‚T1‘

---

<sup>46</sup> Die Varianzkomponente  $u_{1j}$  ist eingeklammert, weil sie in diesem und auch im nächsten Modell fixiert wird. Eine ausführliche Begründung dieses Vorgehens erfolgt im Anschluss an die Beschreibung des Implementationsmodells.

Schließlich wird das Zielmodell, das sogenannte ‚Implementationsmodell (T1 & ImpS)‘ aufgestellt, das zur Beantwortung der Frage nach der Wirksamkeit in Abhängigkeit von der Implementation herangezogen werden soll. Hier wird jetzt neben der Kovariaten ‚T1‘ die wahrgenommene *Implementation* seitens der Schüler, gekennzeichnet mit ‚ImpS‘, ebenfalls zentriert um die Gruppenmittelwerte, auf der ersten Ebene aufgenommen. Die Variable ‚ImpS‘ ist in der Modellbeschreibung als ein Platzhalter für die modulspezifischen Implementationsindikatoren zu verstehen. Diese werden später in Ergebnismodellen mit ‚ImpS.kk‘ für die Implementationswahrnehmung des Moduls Klassenklima, mit ‚ImpS.tp‘ für das Transparenz-Modul und ‚ImpS.llr‘ für das Modul Lern- und Leistungsraum bezeichnet.

### Implementationsmodell (T1 & ImpS)

#### Level 1: Schülerebene

$$Y_{ij} = \beta_{0j} + \beta_{1j}T1 + \beta_{2j}ImpS + r_{ij}$$

$Y_{ij}$  Wert eines Schülers i in Klasse j zum zweiten Messzeitpunkt (abhängige Variable)

$\beta_{0j}$  erwarteter Wert für die Schüler der Klasse j zum zweiten Messzeitpunkt, deren Ausprägungen in den Variablen ‚T1‘ und ‚ImpS‘ dem Klassenmittelwert entsprechen (klassenspezifischer Achsenabschnitt)

$\beta_{1j}$  Einfluss (Kontrolle) der Variablen ‚T1‘ auf die Werte zum zweiten Messzeitpunkt für Schüler der Klasse j (klassenspezifischer Steigungskoeffizient ‚T1‘)

$\beta_{2j}$  Einfluss der Variablen ‚ImpS‘ auf die Werte zum zweiten Messzeitpunkt für Schüler der Klasse j (klassenspezifischer Steigungskoeffizient ‚ImpS‘)

$r_{ij}$  Abweichung des beobachteten Werts eines Schülers i in Klasse j von seinem vorhergesagten Wert zum zweiten Messzeitpunkt

#### Level 2: Klassenebene

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + (u_{1j})$$

$$\beta_{2j} = \gamma_{20} + (u_{2j})^{47}$$

$\gamma_{00}$  erwarteter Wert für die Schüler aller Klassen zum zweiten Messzeitpunkt, deren Ausprägungen in den Variablen ‚T1‘ und ‚ImpS‘ ihren jeweiligen Klassenmittelwerten entsprechen (mittlerer Achsenabschnitt)

<sup>47</sup> Die Varianzkomponente  $u_{2j}$  ist eingeklammert, weil sie in diesem Modell fixiert wird. Eine ausführliche Begründung dieses Vorgehens erfolgt im Anschluss an die Beschreibung des Implementationsmodells.

$\gamma_{10}$	mittlerer Einfluss (Kontrolle) der Variablen ‚T1‘ auf die Werte zum zweiten Messzeitpunkt für Schüler aller Klassen (mittlerer Steigungskoeffizient ‚T1‘)
$\gamma_{20}$	mittlerer Einfluss (Kontrolle) der Variablen ‚ImpS‘ auf die Werte zum zweiten Messzeitpunkt für Schüler aller Klassen (mittlerer Steigungskoeffizient ‚ImpS‘)
$u_{0j}$	Abweichung der Klasse j vom mittleren Achsenabschnitt
$u_{1j}$	Abweichung der Klasse j vom mittleren Steigungskoeffizienten ‚T1‘
$u_{2j}$	Abweichung der Klasse j vom mittleren Steigungskoeffizienten ‚ImpS‘

Auch für dieses Modell wird zur Verbesserung des Verständnisses der einzelnen Modellparameter die Gesamtgleichung ohne Zufallseffekte gebildet.

#### **Gesamtgleichung Implementationsmodell (T1 & ImpS) ohne Zufallseffekte**

$$Y_{ij} = \gamma_{00} + \gamma_{10}T1 + \gamma_{20}ImpS$$

Aufgrund der Zentrierung der Variablen um ihren Gruppenmittelwert stellt  $\gamma_{00}$  den erwarteten Wert zum zweiten Messzeitpunkt für solche Schüler dar, deren Werte in den Variablen ‚T1‘ und ‚ImpS‘ ihrem Klassenmittelwert entsprechen. Der für die Fragestellung interessierende Parameter  $\gamma_{20}$  beschreibt den mittleren Einfluss der modulspezifischen Implementation auf das Kriterium unter Kontrolle der Ausgangswerte ( $\gamma_{10}$ ).

Das hier dargestellte Implementationsmodell repräsentiert beispielhaft ein Modell mit einem Implementationsindikator. Bezogen auf das erste Projektjahr stehen jedoch für die drei Module Klassenklima, Transparenz und Lern- und Leistungsraum die modulspezifischen Implementationsprädiktoren ‚ImpS.kk‘, ‚ImpS.tp‘ und ‚ImpS.llr‘ zur Verfügung. Für diese sind Implementationsmodelle mit einer unterschiedlichen Kombination der Prädiktoren von zwei bis hin zu allen drei Prädiktoren denkbar. Die Gesamtgleichungen ohne Zufallseffekte für ein Modell mit zwei und drei Implementationsprädiktoren werden nachfolgend beschrieben:

#### **Gesamtgleichung Implementationsmodell (T1, ImpS.kk & ImpS.tp) ohne Zufallseffekte**

$$Y_{ij} = \gamma_{00} + \gamma_{10}T1 + \gamma_{20}ImpS.kk + \gamma_{30}ImpS.tp$$

#### **Gesamtgleichung Implementationsmodell (T1, ImpS.kk, ImpS.tp & ImpS.llr) ohne Zufallseffekte**

$$Y_{ij} = \gamma_{00} + \gamma_{10}T1 + \gamma_{20}ImpS.kk + \gamma_{30}ImpS.tp + \gamma_{40}ImpS.llr$$

Die neu hinzugekommenen Parameter  $\gamma_{30}$  und  $\gamma_{40}$  beschreiben wie  $\gamma_{20}$  den mittleren Einfluss der Implementation des jeweiligen Moduls unter Kontrolle der anderen Variablen. Für die Gesamtgleichung mit zwei Implementationsprädiktoren sind auch alle anderen denkbaren Zweier-Kombinationen der drei Prädiktoren möglich. Zur Beantwortung der



Forschungsfrage wird das Modell mit signifikanten Prädiktoren und der besten Modellgüte als das finale Modell herangezogen. Bei der Darstellung der Ergebnisse werden die Modelle berichtet, die folgendem Vorgehen bei der Modellbildung entsprechen: Zunächst wird die zeitlich früheste Maßnahme – der Implementationsindikator Klassenklima – in das Modell aufgenommen. Wird dieser Prädiktor signifikant und weisen die Kriterien zur Modellgüte im Vergleich zum autoregressiven Modell auf eine bessere Passung des Modells zu den Daten hin, wird der Prädiktor beibehalten und der chronologisch nächste Implementationsprädiktor Transparenz wird zusätzlich in das Modell aufgenommen. Ist jedoch der Implementationsindikator Klassenklima nicht signifikant und passt das Modell laut der Kriterien zur Modellgüte nicht besser als das autoregressive Modell, wird der Klassenklima-Prädiktor aus dem Modell entfernt und in einem nächsten Modell nur der Indikator zur Transparenz aufgenommen. Dieser wird wiederum nur dann in einem weiteren Modell beibehalten, wenn er signifikant wird und zur Verbesserung der Modellgüte beiträgt. Ebenso wird auch mit dem Implementationsprädiktor Lern- und Leistungsraum verfahren. Der  $\chi^2$ -Test als ein Kriterium zur Beurteilung der Modellgüte und auch das Pseudo- $R^2$  werden im Vergleich zu den jeweils benachbarten, geschachtelten Modellen ermittelt: bei Modellen mit einem Prädiktor im Vergleich zum autoregressiven Modell, bei Modellen mit zwei Prädiktoren im Vergleich zum Modell mit nur einem Prädiktor und bei einem Modell mit allen drei Prädiktoren im Vergleich zum Vorgängermodell mit zwei Prädiktoren. Ein Modell mit allen drei Implementationsprädiktoren, wie in der zuletzt beschriebenen Gesamtgleichung, kommt in dieser Arbeit nicht vor, da bei keinem Kriterium alle drei Prädiktoren signifikant werden und die vollständigen Modelle nie die besten Indizes zur Modellgüte aufweisen.<sup>48</sup>

Zur Analyse der Frage nach der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler wird der Einfluss der wahrgenommenen modulspezifischen Implementation als ein Level-1-Merkmal modelliert. Entsprechend interessiert vor allem die Aufklärung der Schüler-Varianz (Pseudo- $R^2$  von  $r_{ij}$ ). Diese kann nur ermittelt werden, wenn die Varianz  $r_{ij}$  auch modelliert wird. Auf der zweiten Klassenebene werden keine Prädiktoren aufgenommen, die Varianz zwischen den Klassen ist für die Fragestellung folglich nicht von Interesse und wird durch die Zentrierung um die Klassenmittelwer-

---

<sup>48</sup> Zur Kontrolle von möglichen Supressoreffekten wurden auch Modelle mit Kombinationen von nicht signifikanten Prädiktoren getestet. Eine genaue Beschreibung von Supressoreffekten geben Eid et al. (2013). Da keines dieser „Supressor-Modelle“ zu einer besseren Modellgüte beiträgt, werden jedoch nur die Modelle berichtet, die dem oben beschriebenen Vorgehen entsprechen.

te (,Group-Mean-Centering‘) nivelliert. Deshalb werden die entsprechenden eingeklammerten Varianzkomponenten  $u_{1j}$  und  $u_{2j}$ , die sich auf die Steigungskoeffizienten beziehen, in den kovarianzanalytischen Veränderungsmodellen fixiert. Die Varianz zwischen den Klassen ( $u_{0j}$ ) bezogen auf den Achsenabschnitt wird aus methodischen Gründen nicht fixiert, da mindestens eine Varianzkomponente pro Ebene modelliert werden muss, um die Iteration und Konvergenz der Modelle in HLM 6.08 zu ermöglichen.

#### **6.5.5 Zusammenfassung: Methodisches Vorgehen – Mehrebenenanalysen**

In dem Unterkapitel 6.5 wurde das methodische Vorgehen begründet und erläutert. Das in vorliegenden Arbeit gewählte Auswertungsverfahren, die Mehrebenenanalyse, ist für die FOSIS2-Daten besonders geeignet: Durch den Einsatz dieser Methode können sowohl die kontextuelle Hierarchisierung (Schüler sind in Klassen geschachtelt) als auch die aus den Mehrfachmessungen entstandenen Abhängigkeiten der Daten berücksichtigt werden. Für die Beantwortung der Frage nach der Wirksamkeit des FOSIS2-Programms werden lineare Wachstumskurvenmodelle und diskontinuierliche Veränderungsmodelle durchgeführt. Um den Einfluss der Implementation auf die Zielkriterien zu testen, werden der Kovarianzanalyse ähnelnde Veränderungsmodelle aufgestellt. Im anschließenden Kapitel 7 erfolgt die Darstellung der Ergebnisse dieser Mehrebenenmodelle.

## 7 Ergebnisse

Nachfolgend werden die Ergebnisse der unter 6.5.3 und 6.5.4 beschriebenen Modelle für jedes motivational-relevante Zielkriterium erläutert. Zur Evaluation der Wirksamkeit des Programms werden zunächst die Ergebnisse der linearen und diskontinuierlichen Modelle beschrieben. Im Anschluss daran werden zur Analyse der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler die Ergebnisse der kovarianzanalytischen Veränderungsmodelle berichtet. Für jede Art der Modellierung (linear, diskontinuierlich und kovarianzanalytisch) wird jeweils die erste Ergebnis-Tabelle vollständig und ausführlich beschrieben, um so exemplarisch die Lesart der Tabelle zu erläutern. Zur Reduktion der Darstellung und zur Vereinfachung der Lesbarkeit werden bei den anschließenden Tabellen nur diejenigen Koeffizienten und Parameter der Zielmodelle beschrieben, die zur Beantwortung der Fragestellungen herangezogen werden. Die nach den Kriterien zur Beurteilung der Modellgüte jeweils besten Modelle sind in der Kopfzeile jeder Tabelle fett markiert und die ineinander geschachtelten Modelle werden mit einer jeweils höheren Nummer gekennzeichnet (z. B. 1., 2., 3.). Zur Veranschaulichung der Ergebnisse werden die lineare und die diskontinuierliche Veränderung beispielhaft für die ersten beiden beschriebenen Kriterien, die schulische Selbstwirksamkeit und die Prüfungsängstlichkeit, grafisch dargestellt. Um den Umfang des Ergebnis-Kapitels zu reduzieren, finden sich die Grafiken der übrigen Kriterien im Anhang E.

### 7.1 Wirksamkeitsevaluation

Die Wirksamkeit des FOSIS2-Programms im Hinblick auf die motivationalen Variablen wird methodisch zum einen im Rahmen linearer Wachstumskurvenmodelle und zum anderen im Rahmen diskontinuierlicher Veränderungsmodelle evaluiert. Erstere dienen der Überprüfung der gruppenspezifischen linearen Veränderung über beide Projektjahre hinweg. Letztere sollen jahresspezifische Unterschiede zwischen der Interventions- und der Kontrollgruppe aufdecken. Eine Übersicht über die deskriptiven Statistiken der Kriterien findet sich im Anhang D.

### 7.1.1 Lineare Wachstumskurvenmodelle

Die Modellbildung und die jeweiligen Modellparameter der linearen Wachstumskurvenmodelle wurden ausführlich in Abschnitt 6.5.3 erläutert.

#### *Schulische Selbstwirksamkeitserwartung*

Tabelle 6 gibt Auskunft über die lineare Veränderung in der schulischen Selbstwirksamkeitserwartung über die Zeit. Die Kriterien zur Beurteilung der Modellgüte beschreiben das Nullmodell als das beste Modell. Zwar fallen die Devianzen des Veränderungs- (1984.966) und des Evaluationsmodells (1983.689) kleiner aus als die Devianz des Nullmodells (1989.443). Jedoch werden die jeweiligen Differenzen der Devianzen (4.477; 1.277) im  $\chi^2$ -Test nicht signifikant. Zudem ist der AIC des Veränderungsmodells mit 2002.966 größer als der des Nullmodells (1997.443) und der AIC des Evaluationsmodells (2005.689) ist sogar am höchsten. Veränderungs- und Evaluationsmodell sind somit nicht eindeutig besser zur Vorhersage der Daten geeignet als das Nullmodell. Nachfolgend soll die Bedeutung der einzelnen Parameter der Modelle erläutert werden.

Das Nullmodell beschreibt noch keine Veränderung, sondern nur die Variation der schulischen Selbstwirksamkeitserwartung über alle Messzeitpunkte und Gruppen hinweg ( $\gamma_{000} = 2.879^{***}$ ). Mithilfe dieses Modells sollen vor allem substantielle Varianzen und die Intraklassen-Korrelation (ICC) ermittelt werden. Im Gegensatz zur Varianz zwischen den Schülern ( $r_{0ij} = 0.168^{***}$ ) fällt die Varianz zwischen den Klassen  $u_{00j}$  im Nullmodell zwar signifikant, jedoch mit 0.008\*\* sehr klein aus. Entsprechend ist auch die Intraklassen-Korrelation (ICC; siehe Tabellenanmerkungen) der Klassenebene, die den Anteil der Varianz zwischen den Klassen an der Gesamtvarianz beschreibt, mit 0.025 ebenfalls als klein zu bewerten (Hox, 2010).

Im zweiten Modell, dem Veränderungsmodell, wird durch die Aufnahme des Level-1-Prädiktors ‚Zeit‘ die lineare Veränderung über beide Projektjahre hinweg modelliert. Es werden noch keine Unterschiede zwischen der Interventions- und der Kontrollgruppe ermittelt. In diesem Modell gibt der Achsenabschnitt  $\gamma_{000}$  (2.858\*\*\*) den Mittelwert aller Klassen zum ersten Messzeitpunkt an. Die zugehörigen Varianzkomponenten  $r_{0ij}$  (0.170\*\*\*) und  $u_{00j}$  (0.012\*\*) werden signifikant und fallen wider Erwarten größer aus als im Nullmodell. Entsprechend wird das Pseudo- $R^2$  der Varianzkomponente  $u_{00j}$  negativ und beträgt -0.500. Ein negatives Pseudo- $R^2$  kann nach Snijders und Bosker (1994) unter bestimmten Umständen vorkommen. Für dieses Problem gibt es derzeit noch keine gute

Lösungsstrategie (Hox, 2010) und ein solches negatives Pseudo- $R^2$  sollte nicht interpretiert werden (Singer & Willett, 2003; eine ausführlichere Darstellung des Problems findet sich im Anhang C2). Der Parameter  $\gamma_{100}$  (0.018) beschreibt die lineare mittlere Veränderung aller Klassen über die Zeit. Diese ist nicht signifikant und die zugehörigen Varianzkomponenten  $r_{1ij}$  (0.009\*) und  $u_{10j}$  (0.0004) fallen sehr klein aus. Für die Frage nach der Wirksamkeit des Programms ist die zuletzt genannte Varianz zwischen den Klassen ( $u_{10j}$ ) von besonderem Interesse. Denn diese soll im Zielmodell 3, dem Evaluationsmodell, durch die Aufnahme des Prädiktors ‚Gruppe‘ auf der dritten Ebene aufgeklärt werden. Da die Komponente  $u_{10j}$  gegen Null geht, liefert sie bereits einen Hinweis auf einen nicht vorhandenen Effekt des Prädiktors ‚Gruppe‘. Denn wenn die Klassen in ihrer Veränderung nicht bedeutsam variieren, kann die Zugehörigkeit zur Interventions- oder Kontrollgruppe im Evaluationsmodell auch keine Varianz bezogen auf die Veränderung aufklären. Genau dies wird mit dem Evaluationsmodell geprüft:

Durch die Aufnahme des Prädiktors ‚Gruppe‘ auf der Klassenebene wird die Veränderung der Interventionsgruppe über die Zeit im Vergleich zur Kontrollgruppe untersucht. Da ein Prädiktor einer höheren Ebene zu keiner Aufklärung der Varianz einer niedrigeren Ebene beitragen kann (Hox, 2010) und keine Prädiktoren auf der Schülerebene aufgenommen werden, verändern sich die Varianzkomponenten auf dieser Ebene  $r_{0ij}$  (0.170\*\*\*) und  $r_{1ij}$  (0.009\*) vom Veränderungs- zum Evaluationsmodell in der Regel nicht. Der Koeffizient  $\gamma_{000}$  (2.824\*\*\*) bildet in diesem Modell den mittleren Wert aller Kontrollklassen zum ersten Messzeitpunkt ab und  $\gamma_{001}$  erfasst die entsprechende Abweichung der Interventionsklassen, also den Unterschied in den Ausgangswerten. Dieser beträgt 0.068 und ist nicht signifikant. Die Varianzkomponente  $u_{00j}$  (0.012\*\*) verändert sich vom Veränderungs- zum Evaluationsmodell nicht und die Aufnahme der Variable ‚Gruppe‘ als Level-3-Prädiktor trägt somit im Vergleich der beiden Modelle<sup>49</sup> zu keiner Aufklärung der Varianz  $u_{00j}$  bei (Pseudo- $R^2$  von  $u_{00j} = 0.000$ ). Der Parameter  $\gamma_{100}$  (0.021) beschreibt die durchschnittliche Veränderung aller Kontrollklassen über die Zeit: Die Kontrollklassen verändern sich nicht signifikant. Den Unterschied zwischen Interventions- und Kontrollklassen in ihrer linearen mittleren Veränderung über die Zeit bildet der Koeffizient  $\gamma_{101}$  (-0.005) ab. Auch dieser sehr kleine Effekt wird nicht signifikant und verdeutlicht somit, dass sich

---

<sup>49</sup> Das Pseudo- $R^2$  wird, wie im methodischen Vorgehen (Anhang C2) beschrieben, immer aus dem Vergleich der jeweiligen Varianzkomponenten aus zwei benachbarten Modellen ermittelt und kann somit auch immer nur im Vergleich zum Nachbarmodell interpretiert werden.

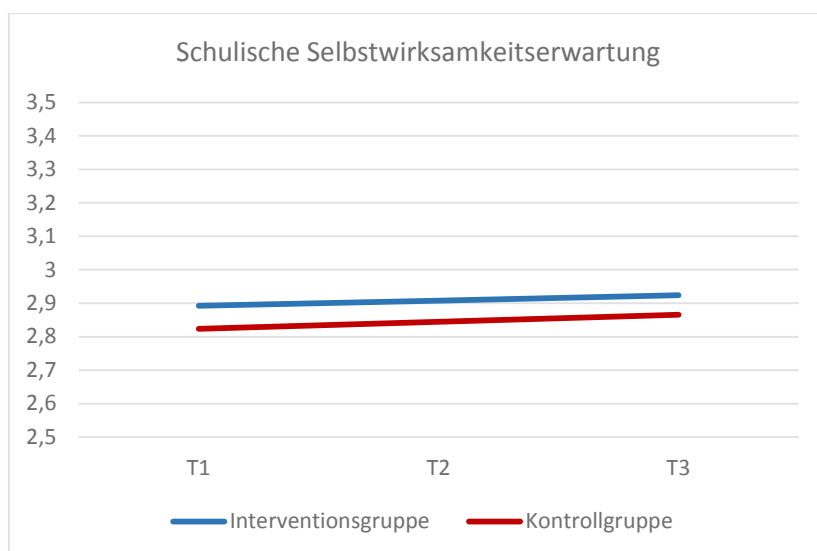
Interventions- und Kontrollklassen nicht bedeutsam in ihrer mittleren Veränderung über die Zeit unterscheiden. Die zugehörige Varianz  $u_{10j}$  (0.0004) zwischen den Klassen, bezogen auf die Steigung, verändert sich vom Veränderungs- zum Evaluationsmodell nicht. Folglich leistet der Prädiktor ‚Gruppe‘ keinen Beitrag zur Aufklärung der Varianzkomponente  $u_{10j}$  und das Pseudo- $R^2$  dieser Varianz beträgt Null.

Zusammenfassend zeigen sich bei ähnlichen Ausgangswerten zwischen Interventions- und Kontrollgruppe keine bedeutsamen Unterschiede hinsichtlich ihrer Entwicklung in der schulischen Selbstwirksamkeitserwartung über die Zeit. Abbildung 11 veranschaulicht dies grafisch: Die gruppenspezifische lineare Veränderung in der schulischen Selbstwirksamkeitserwartung verläuft nahezu parallel in Interventions- (blaue Linie) und Kontrollgruppe (rote Linie). Entgegen der Hypothese verbessert sich die Interventionsgruppe verglichen mit der Kontrollgruppe folglich nicht in ihrer schulischen Selbstwirksamkeitserwartung. Die Kriterien zur Beurteilung der Modellgüte beschreiben das Nullmodell als das beste Modell zur Vorhersage der Daten.

**Tabelle 6:** Lineare Veränderung in der schulischen Selbstwirksamkeitserwartung über beide Projektjahre

n L1 = 1355		1.	2.	3.
n L2 = 566	Parameter	Nullmodell	Veränderungsmodell (Zeit)	Evaluationsmodell (Zeit & Gruppe)
n L3 = 22				
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.879*** (0.03)	2.858*** (0.03)	2.824*** (0.06)
Veränderung über die Zeit	$\gamma_{100}$	--	0.018 (0.01)	0.021 (0.02)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	0.068 (0.07)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	-0.005 (0.03)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.147	0.137	0.137
Level 2:	$r_{0ij}$	0.168***	0.170***	0.170***
	$r_{1ij}$	--	0.009*	0.009*
Level 3:	$u_{00j}$	0.008**	0.012**	0.012**
	$u_{10j}$	--	0.0004	0.0004
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	-0.500	0.000
	$u_{10j}$	--	--	0.000
<b>Modellfit</b>				
Devianz(q)		1989.443(4)	1984.966(9)	1983.689(11)
$\chi^2$ -Differenz		--	4.477	1.277
AIC		1997.443	2002.966	2005.689

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; ICC für Level 3: 0.025



**Abbildung 11: Lineare Veränderung in der schulischen Selbstwirksamkeitserwartung über beide Projektjahre (T1-T3 = Messzeitpunkt 1-3)**

### ***Prüfungsängstlichkeit***

Tabelle 7 gibt Auskunft über die lineare Veränderung der Prüfungsängstlichkeit über die Zeit. Die ICC (siehe Tabellenanmerkungen) fällt mit 0.043 eher klein aus (Hox, 2010). Nach den Kriterien zur Beurteilung der Modellgüte verfügt das Evaluationsmodell über den besten Modellfit. Verglichen mit den entsprechenden Kennwerten des Null- (Devianz = 2761.368; AIC = 2769.368) und des Veränderungsmodells (Devianz = 2729.233; AIC = 2747.233) sind sowohl die Devianz (2722.148) als auch der AIC (2744.148) des Evaluationsmodells am kleinsten. Zudem wird die Differenz zwischen dem Veränderungs- und Evaluationsmodell (7.085\*) im  $\chi^2$ -Test signifikant und weist somit ebenfalls auf die Güte des Evaluationsmodells hin. Nachfolgend werden die Koeffizienten dieses besten Modells näher erläutert:

Die Ausgangswerte zwischen Interventions- und Kontrollgruppe unterscheiden sich nicht signifikant ( $\gamma_{001} = -0.101$ ). Dennoch trägt die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ zu einer Varianzaufklärung von 10 % der Gruppenunterschiede in den Ausgangswerten bei (Pseudo- $R^2$  von  $u_{00j} = 0.100$ ). Die Varianzkomponente  $u_{00j}$  (0.018\*\*) des Evaluationsmodells bleibt verglichen mit dem Veränderungsmodell signifikant. Trotz der Aufnahme des Prädiktors ‚Gruppe‘ besteht folglich weiterhin eine substantielle Varianz zwischen den Klassen bezogen auf die Ausgangswerte, die jedoch nicht auf die Zugehörigkeit zur Interventions- oder Kontrollgruppe zurückgeführt werden kann. Der Effekt  $\gamma_{101}$  (-0.078) beantwortet die Frage nach der gruppenspezifischen Veränderung über die Zeit. Dieser besagt, dass sich die Prüfungsängstlichkeit bei den Interventionsschülern im Verlauf der bei-

den Projektjahre stärker reduziert als bei den Kontrollschülern. Das heißt, die Kontrollschüler berichten ebenfalls eine Abnahme in der Prüfungsängstlichkeit über die Zeit ( $\gamma_{100} = -0.060^*$ ), in der Interventionsgruppe ist die Reduktion jedoch pro Projektjahr um 0.078 Skaleneinheiten stärker als in der Kontrollgruppe. Dieser Gruppenunterschied in der Veränderung ist zwar nicht signifikant, die Zugehörigkeit zur Gruppe erklärt aber verglichen mit dem Veränderungsmodell 40 % der Varianz auf (Pseudo- $R^2$  von  $u_{10j} = 0.400$ ) und die zugehörige Varianzkomponente  $u_{10j}$  (0.003) ist im Evaluationsmodell nicht mehr signifikant. Die Varianzkomponenten der Schülerebene verändern sich vom Veränderungs- zum Evaluationsmodell bis auf minimale, durch die Schätzverfahren verursachte Abweichungen nicht, da im Zielmodell keine zusätzlichen Prädiktoren auf der Schülerebene aufgenommen werden. Dies gilt auch für die Modelle der nachfolgenden Kriterien.

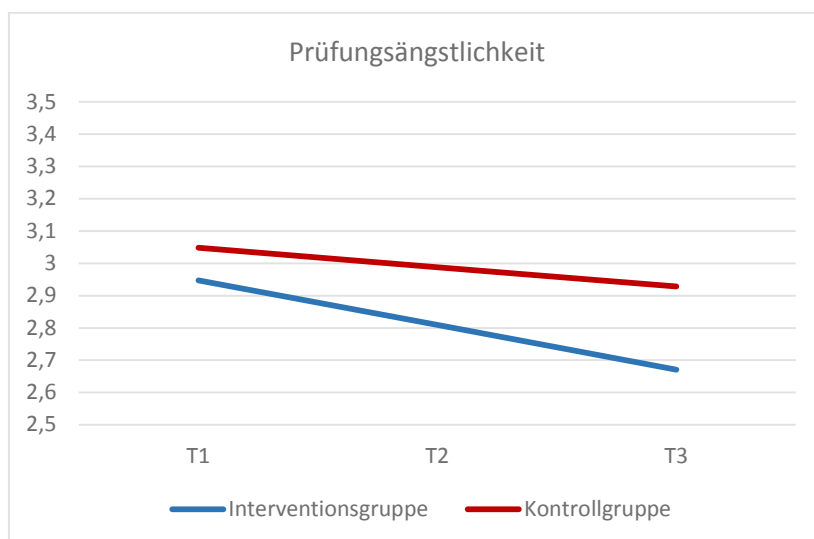
Zusammenfassend lässt sich festhalten: Bei ähnlichen Ausgangswerten reduziert sich die Prüfungsängstlichkeit in den Interventionsklassen im Verlauf der beiden Projektjahre zwar stärker als in den Kontrollklassen (siehe auch Abbildung 12: Interventionsgruppe = blaue Linie; Kontrollgruppe = rote Linie). Dieser Unterschied in der Veränderung ist jedoch nicht signifikant. Gleichwohl klärt die Gruppenzugehörigkeit 40 % der Varianz bezogen auf die Veränderung auf und die Kriterien zur Beurteilung der Modellgüte beschreiben das Evaluationsmodell als das am besten geeignete Modell zur Vorhersage der Daten.



**Tabelle 7:** Lineare Veränderung in der Prüfungsängstlichkeit über beide Projektjahre

n L1 = 1339 n L2 = 566 n L3 = 22	Parameter	1. Nullmodell	2. Veränderungs- modell (Zeit)	3. <b>Evaluationsmodell (Zeit &amp; Gruppe)</b>
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.898*** (0.04)	2.999*** (0.04)	3.048*** (0.08)
Veränderung über die Zeit	$\gamma_{100}$	--	-0.096** (0.02)	-0.060* (0.03)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.101 (0.08)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	-0.078 (0.05)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.303	0.286	0.286
Level 2:	$r_{0ij}$	0.211***	0.176***	0.176***
	$r_{1ij}$	--	0.003	0.003
Level 3:	$u_{00j}$	0.023***	0.020**	0.018**
	$u_{10j}$	--	0.005*	0.003
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.130	0.100
	$u_{10j}$	--	--	0.400
<b>Modellfit</b>				
Devianz(q)		2761.368(4)	2729.233(9)	2722.148(11)
$\chi^2$ -Differenz		--	32.135***	7.085*
AIC		2769.368	2747.233	2744.148

Anmerkungen n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.043



**Abbildung 12:** Lineare Veränderung in der Prüfungsängstlichkeit über beide Projektjahre (T1-T3 = Messzeitpunkt 1-3)

### ***Schulbezogenes Stresserleben – wahrgenommene Herausforderung***

Tabelle 8 beschreibt die lineare Veränderung in der wahrgenommenen Herausforderung über die Zeit. Mit 0.05 ist die ICC (siehe Tabellenanmerkungen) nach Hox (2010) als klein, aber bedeutsam für das Verfahren zu bewerten. Die Kriterien zur Beurteilung der Modellgüte weisen das Evaluationsmodell als das beste Modell aus. Im Vergleich zu den Gütekriterien des Null- (Devianz = 2777.068; AIC = 2785.068) und des Veränderungsmodells (Devianz = 2709.500; AIC = 2727.500) fallen sowohl die Devianz (2692.773) als auch der AIC (2714.773) des Evaluationsmodells am geringsten aus. Dies spricht für die Güte des Modells. Die Differenz zwischen dem Veränderungs- und Evaluationsmodell (16.727\*\*\*) wird im  $\chi^2$ -Test zudem signifikant und ist somit ein weiteres Indiz für die Güte des Evaluationsmodells.

Anders als bei den zuvor beschriebenen Konstrukten unterscheiden sich die Ausgangswerte zwischen Interventions- und Kontrollgruppe signifikant ( $\gamma_{001} = 0.303^{***}$ ). Verglichen mit der entsprechenden Varianzkomponente des Veränderungsmodells trägt die Aufnahme des Prädiktors Gruppe zu einer Aufklärung von 72 % bei (Pseudo- $R^2$  von  $u_{00j} = 0.720$ ) und die Varianz  $u_{00j}$  (0.007) ist im Evaluationsmodell nicht mehr signifikant. Die wahrgenommene Herausforderung liegt folglich in der Interventionsgruppe zu Beginn des Projekts signifikant höher als in der Kontrollgruppe und der Unterschied kann durch die Zugehörigkeit zur Gruppe zu 72 % aufgeklärt werden. Im Hinblick auf die Veränderung über die Zeit zeigt sich entgegen der Erwartung sowohl in der Kontroll- als auch in der Interventionsgruppe ein Abfall im wahrgenommenen Herausforderungserleben. Denn die beiden Steigungskoeffizienten, die zum einen die mittlere lineare Veränderung in den Kontrollklassen beschreiben ( $\gamma_{100} = -0.083^*$ ) und zum anderen den Unterschied in der linearen mittleren Veränderung zwischen Interventions- und Kontrollklassen erfassen ( $\gamma_{101} = -0.108^*$ ), weisen beide ein negatives Vorzeichen auf. In den Interventionsklassen reduziert sich die wahrgenommene Herausforderung pro Projektjahr sogar signifikant, um 0.108 mehr Skaleneinheiten als in den Kontrollklassen. Der Prädiktor ‚Gruppe‘ klärt vom Veränderungs- zum Evaluationsmodell 33 % der Varianz bezogen auf diesen Gruppenunterschied in der Veränderung auf (Pseudo- $R^2$  von  $u_{10j} = 0.333$ ).

Zusammenfassend zeigt sich bei den Interventionsschülern im Verlauf der beiden Projektjahre entgegen der Erwartung eine signifikant stärkere Abnahme der wahrgenommenen Herausforderung als bei den Kontrollschülern. Bei der später erfolgenden Diskussion die-

ses Ergebnisses müssen die signifikant unterschiedlichen Ausgangswerte beachtet werden. Im Hinblick auf den Modellfit passt das Evaluationsmodell am besten zu den Daten.

**Tabelle 8:** Lineare Veränderung in der wahrgenommenen Herausforderung über beide Projektjahre

n L1 = 1352 n L2 = 566 n L3 = 22		1.	2.	3.
	Parameter	Nullmodell	Veränderungsmodell (Zeit)	Evaluationsmodell (Zeit & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.382***(0.04)	2.524***(0.05)	2.372***(0.04)
Veränderung über die Zeit	$\gamma_{100}$	--	-0.137***(0.02)	-0.083*(0.03)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	0.303***(0.07)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	-0.108*(0.04)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.298	0.243	0.243
Level 2:	$r_{0ij}$	0.215***	0.214***	0.214***
	$r_{1ij}$	--	0.037***	0.037***
Level 3:	$u_{00j}$	0.027***	0.025***	0.007
	$u_{10j}$	--	0.003	0.002
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$		0.074	0.720
	$u_{10j}$			0.333
<b>Modellfit</b>				
Devianz(q)		2777.068(4)	2709.500(9)	2692.773(11)
$\chi^2$ -Differenz			67.568***	16.727***
AIC		2785.068	2727.500	2714.773

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.05

### **Schulbezogenes Stresserleben – wahrgenommene Bedrohung**

In Tabelle 9 ist die lineare Veränderung der wahrgenommenen Bedrohung über die Zeit dargestellt. Die ICC (siehe Tabellenanmerkungen) fällt mit 0.043 eher klein aus (Hox, 2010). Nach den Kriterien zur Beurteilung der Modellgüte weist das Evaluationsmodell den besten Modellfit auf. Sowohl die Devianz (2681.449) als auch der AIC (2703.449) des Evaluationsmodells sind verglichen mit den entsprechenden Koeffizienten des Null- (Devianz = 2738.082; AIC = 2746.082) und des Veränderungsmodells (Devianz = 2690.465; AIC = 2708.465) am geringsten. Auch der  $\chi^2$ -Test weist auf die Güte des Evaluationsmodells hin, denn die Differenz zwischen dem Veränderungs- und Evaluationsmodell (9.016\*) wird signifikant.

Die Ausgangswerte zwischen Interventions- und Kontrollgruppe unterscheiden sich nicht signifikant ( $\gamma_{001} = -0.052$ ) und die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ führt vom Veränderungs- zum Evaluationsmodell zu einer Varianzaufklärung von 6.7 % der Grup-

penunterschiede in den Ausgangswerten (Pseudo- $R^2$  von  $u_{00j} = 0.067$ ). Wie bereits bei der Prüfungsängstlichkeit bleibt die Varianzkomponente  $u_{00j}$  ( $0.014^{**}$ ) des Evaluationsmodells verglichen mit dem Veränderungsmodell signifikant. Demnach besteht weiterhin eine substantielle Variation zwischen den Klassen bezogen auf die Ausgangswerte, die durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe nicht aufgeklärt werden kann. Der Effekt  $\gamma_{101}$  ( $-0.125^*$ ) ist negativ und wird signifikant. Im Vergleich zu den Kontrollschülern, deren Bedrohungserleben sich nicht bedeutsam verändert ( $\gamma_{100} = -0.007$ ), nehmen die Interventionsschüler im Verlauf der beiden Projektjahre folglich signifikant weniger Bedrohung wahr. Durch die Aufnahme des Prädiktors ‚Gruppe‘ wird vom Veränderungs- zum Evaluationsmodell 50% der Varianz bezogen auf die Veränderung erklärt (Pseudo- $R^2$  von  $u_{10j} = 0.500$ ), wobei die Varianzkomponente  $u_{10j}$  ( $0.004^*$ ) im Evaluationsmodell noch immer signifikant ist. Dies bedeutet, bezogen auf die Veränderung über die Zeit existiert weiterhin eine substantielle Variation zwischen den Klassen, die durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe nicht aufgeklärt werden kann.

Zusammenfassend lässt sich festhalten: Bei ähnlichen Ausgangswerten reduziert sich die wahrgenommene Bedrohung in den Interventionsklassen signifikant stärker als in den Kontrollklassen. Nach den Kriterien zur Beurteilung der Modellgüte weist das Evaluationsmodell den besten Modellfit auf.

**Tabelle 9:** Lineare Veränderung in der wahrgenommenen Bedrohung über beide Projektjahre

n L1 =1353 n L2 = 566 n L3 = 22	Parameter	1. Nullmodell	2. Veränderungs- modell (Zeit)	3. <b>Evaluationsmodell (Zeit &amp; Gruppe)</b>
<b><i>Feste Effekte</i></b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.157*** (0.04)	2.225*** (0.04)	2.251*** (0.06)
Veränderung über die Zeit	$\gamma_{100}$	--	-0.067* (0.03)	-0.007 (0.03)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.052 (0.08)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	-0.125* (0.05)
<b><i>Zufallseffekte (Varianzkomponenten)</i></b>				
Level 1:	$e_{tij}$	0.296	0.262	0.262
Level 2:	$r_{0ij}$	0.195***	0.130***	0.130***
	$r_{1ij}$	--	0.022*	0.022*
Level 3:	$u_{00j}$	0.022***	0.015**	0.014**
	$u_{10j}$	--	0.008**	0.004*
<b><i>Varianzaufklärung (Pseudo-R<sup>2</sup>)</i></b>				
Level 3:	$u_{00j}$	--	0.318	0.067
	$u_{10j}$	--	--	0.500
<b><i>Modellfit</i></b>				
Devianz(q)		2738.082(4)	2690.465(9)	2681.449(11)
$\chi^2$ -Differenz		--	47.617***	9.016*
AIC		2746.082	2708.465	2703.449

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.043

### ***Schulbezogenes Stresserleben – wahrgenommener Verlust***

Tabelle 10 gibt Auskunft über die lineare Veränderung des wahrgenommenen Verlusts. Auch das Nullmodell zur Vorhersage des wahrgenommenen Verlusts weist nach Hox (2010) eine eher kleine ICC (0.022) auf (siehe Tabellenanmerkungen). Das Evaluationsmodell verfügt über den besten Modellfit. Denn sowohl die Devianz (2917.326) als auch der AIC (2939.326) dieses Zielmodells sind geringer als die entsprechenden Kennwerte des Null- (Devianz = 2947.758; AIC = 2955.758) und des Veränderungsmodells (Devianz = 2924.277; AIC = 2942.277). Der  $\chi^2$ -Test bestätigt ebenfalls die Güte des Evaluationsmodells. Denn die Differenz zwischen dem Veränderungs- und Evaluationsmodell (6.951\*\*) ist signifikant.

Wie bei der wahrgenommenen Herausforderung unterscheiden sich die Ausgangswerte zwischen der Interventions- und der Kontrollgruppe signifikant ( $\gamma_{001} = -0.168^{**}$ ) und die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ führt vom Veränderungs- zum Evaluationsmodell zu einer Varianzaufklärung von 75% der Gruppenunterschiede in den Ausgangswerten (Pseudo-R<sup>2</sup> von  $u_{00j} = 0.750$ ). Die Interventionsgruppe nimmt zu Beginn des Pro-

jekts signifikant weniger Verlust wahr als die Kontrollgruppe. Im Hinblick auf die Veränderung über die Zeit berichten beide Gruppen eine Zunahme des schulbezogenen Verlusts ( $\gamma_{100} = 0.061^*$ ;  $\gamma_{101} = 0.011$ ). Der Anstieg des Verlusterlebens ist in der Interventionsgruppe pro Projektjahr sogar um 0.011 Skaleneinheiten höher als in der Kontrollgruppe. Dieser Gruppenunterschied in der Veränderung ist jedoch nicht signifikant. Verglichen mit dem Veränderungsmodell leistet der Prädiktor ‚Gruppe‘ im Evaluationsmodell entsprechend auch keinen Beitrag zur Aufklärung der Varianz, die sich auf die Veränderung bezieht. Die Aufklärungsrate beträgt 0 % (Pseudo- $R^2$  von  $u_{10j} = 0.000$ ).

Zusammenfassend lässt sich festhalten: Hinsichtlich des Verlusterlebens gibt es im Laufe der beiden Projektjahre keine substantiellen Unterschiede zwischen der Interventions- und der Kontrollgruppe. Bei der später erfolgenden Diskussion dieses Ergebnisses müssen die signifikant unterschiedlichen Ausgangswerte berücksichtigt werden. Im Hinblick auf die Modellgüte weist das Evaluationsmodell den besten Modellfit auf.

**Tabelle 10:** Lineare Veränderung im wahrgenommenen Verlust über beide Projektjahre

n L1 = 1353 n L2 = 565 n L3 = 22		1.	2.	3.
Parameter		Nullmodell	Veränderungsmodell (Zeit)	Evaluationsmodell (Zeit & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	1.782***(0.04)	1.712***(0.03)	1.797***(0.04)
Veränderung über die Zeit	$\gamma_{100}$	--	0.068***(0.02)	0.061*(0.03)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.168**(0.05)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	0.011(0.04)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.359	0.320	0.320
Level 2:	$r_{0ij}$	0.208***	0.191***	0.191***
	$r_{1ij}$	--	0.033**	0.033**
Level 3:	$u_{00j}$	0.013**	0.004	0.001
	$u_{10j}$	--	0.002	0.002
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.692	0.750
	$u_{10j}$	--	--	0.000
<b>Modellfit</b>				
Devianz(q)		2947.758(4)	2924.277(9)	2917.326(11)
$\chi^2$ -Differenz		--	23.481***	6.951*
AIC		2955.758	2942.277	2939.326

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; ICC für Level 3: 0.022

### ***Wahrgenommene Binnendifferenzierung***

In Tabelle 11 wird die lineare Veränderung in der wahrgenommenen Binnendifferenzierung im Unterricht über die Zeit beschrieben. Die ICC (siehe Tabellenanmerkungen) ist nach Hox (2010) mit 0.056 als klein, aber durchaus bedeutsam für das Verfahren der Mehrebenenanalyse zu bewerten. Veränderungs- und Evaluationsmodell unterscheiden sich kaum in ihrer Modellgüte. Zwar weist das Evaluationsmodell im Vergleich zum Null- (Devianz = 2622.281; AIC = 2630.281) und Veränderungsmodell (Devianz = 2576.216; AIC = 2594.216) die geringste Devianz (2571.509) und einen geringfügig kleineren AIC (2593.509) auf, was für die Güte dieses Modells spricht. Die Differenz der Devianzen des Veränderungs- und des Evaluationsmodells ist jedoch mit 4.707 sehr klein und nicht signifikant. Der AIC spricht folglich für eine minimal bessere Güte des Evaluationsmodells, der  $\chi^2$ -Test hingegen erklärt das Veränderungsmodell für das bessere Modell. Obwohl das Evaluationsmodell hier nicht eindeutig über den besten Modellfit verfügt, sollen trotzdem die Koeffizienten dieses Modells beschrieben werden. Denn das Evaluationsmodell stellt das Zielmodell dar, das zur Beantwortung der Frage nach der Wirksamkeit des FOSIS2-Programms herangezogen wird.

Hinsichtlich der festen Effekte des Evaluationsmodells ist zunächst festzuhalten, dass sich die Ausgangswerte zwischen Interventions- und Kontrollgruppe nicht signifikant unterscheiden ( $\gamma_{001} = 0.062$ ) und die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ vom Veränderungs- zum Evaluationsmodell nur zu einer Varianzaufklärung von 5.6 % der Gruppenunterschiede in den Ausgangswerten beiträgt (Pseudo- $R^2$  von  $u_{00j} = 0.056$ ). Wie bei der Prüfungsängstlichkeit und der wahrgenommenen Bedrohung bleibt die Varianzkomponente  $u_{00j}$  (0.017\*\*) im Evaluationsmodell verglichen mit dem Veränderungsmodell signifikant. Trotz der Aufnahme des Prädiktors ‚Gruppe‘ besteht demzufolge weiterhin eine substantielle Variation zwischen den Klassen bezogen auf die Ausgangswerte. Diese lässt sich jedoch nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe erklären. Der für die Beantwortung der Frage nach der Wirksamkeit des Programms wichtigste Effekt ist  $\gamma_{101}$  (0.090). Dieser weist durch das positive Vorzeichen zwar in die erwartete Richtung: Die Interventionsklassen nehmen verglichen mit den Kontrollschülern ( $\gamma_{100} = 0.032$ ) im Verlauf der beiden Projektjahre durchschnittlich etwas mehr Binnendifferenzierung im Unterricht wahr. Zudem trägt der Prädiktor ‚Gruppe‘ vom Veränderungs- zum Evaluationsmodell zu einer Varianzaufklärung von 14.3 % bei (Pseudo- $R^2$  von  $u_{10j} = 0.143$ ). Der Gruppenunterschied in der Veränderung ist jedoch nicht signifikant und die Varianzkom-

ponente  $u_{10j}$  (0.012\*\*\*) bleibt im Evaluationsmodell noch immer signifikant. Bezogen auf die Veränderung über die Zeit existiert folglich weiterhin eine substantielle Variation zwischen den Klassen, die nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe aufgeklärt werden kann.

Zusammengefasst nehmen die Interventionsschüler bei ähnlichen Ausgangswerten mehr Binnendifferenzierung im Verlauf der beiden Projektjahre wahr als die Kontrollschüler. Dieser Effekt ist jedoch nicht signifikant und die Kriterien zur Beurteilung der Modellgüte weisen darauf hin, dass sich Veränderungs- und Evaluationsmodell hinsichtlich ihrer Modellgüte nicht eindeutig unterscheiden.

**Tabelle 11:** Lineare Veränderung in der wahrgenommenen Binnendifferenzierung über beide Projektjahre

n L1 = 1341 n L2 = 566 n L3 = 22		1.	2.	3.
Parameter		Nullmodell	Veränderungsmodell (Zeit)	Evaluationsmodell (Zeit & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.136*** (0.04)	2.059*** (0.04)	2.028*** (0.06)
Veränderung über die Zeit	$\gamma_{100}$	--	0.075* (0.03)	0.032 (0.04)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	0.062 (0.08)
Gruppenunterschied in der Veränderung über die Zeit	$\gamma_{101}$	--	--	0.090 (0.06)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.331	0.273	0.273
Level 2:	$r_{0ij}$	0.088***	0.175***	0.175***
	$r_{1ij}$	--	0.040***	0.040***
Level 3:	$u_{00j}$	0.025***	0.018**	0.017**
	$u_{10j}$	--	0.014***	0.012***
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.28	0.056
	$u_{10j}$	--	--	0.143
<b>Modellfit</b>				
Devianz(q)		2622.281(4)	2576.216(9)	2571.509(11)
$\chi^2$ -Differenz		--	46.065***	4.707
AIC		2630.281	2594.216	2593.509

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; ICC für Level 3: 0.056

### 7.1.2 Diskontinuierliche Veränderungsmodelle

Die diskontinuierlichen Veränderungsmodelle dienen der Prüfung jahresspezifischer Unterschiede zwischen der Interventions- und der Kontrollgruppe. Anders als bei den linearen Wachstumskurvenmodellen, wo die Veränderung als linearer Term über beide Projektjahre modelliert wird, prüfen die diskontinuierlichen Modelle die sequentielle phasenspezifische Veränderung pro Projektjahr. Wie bereits bei den linearen Modellen soll die erste Tabelle



zur Erläuterung der Lesart ausführlich beschrieben werden. Im Anschluss werden nur die für die Fragestellung wichtigsten Parameter und Koeffizienten dargestellt. Zu beachten ist, dass die diskontinuierlichen Modelle und die linearen Wachstumskurvenmodelle, wie bei der Modellbildung erläutert (siehe 6.5.3), mit jeweils demselben Nullmodell verglichen werden. Da die ICCs der Nullmodelle bereits beschrieben wurden, wird nachfolgend zur Vermeidung von Wiederholungen auf eine erneute Darstellung der ICCs verzichtet. Zur Veranschaulichung der Ergebnisse wird die diskontinuierliche Veränderung wiederum beispielhaft für die schulische Selbstwirksamkeit und die Prüfungsängstlichkeit grafisch dargestellt. Die entsprechenden Grafiken der übrigen Zielkriterien finden sich im Anhang E.

### ***Schulische Selbstwirksamkeitserwartung***

Tabelle 12 gibt Auskunft über die diskontinuierliche, jahresspezifische Veränderung in der schulischen Selbstwirksamkeitserwartung. Hinsichtlich der Kriterien zur Modellgüte fällt auf, dass zwar die Devianzen des Veränderungs- (1975.646) und des Evaluationsmodells (1973.804) kleiner sind als die Devianz des Nullmodells (1989.443). Die Differenzen der jeweiligen Devianzen (13.797; 1.842) sind jedoch nicht signifikant und die AICs des Veränderungs- (2001.646) und des Evaluationsmodells (2005.804) fallen sogar größer aus als der AIC des Nullmodells (1997.443). Veränderungs- und Evaluationsmodell sind somit nicht eindeutig besser zur Vorhersage der Daten geeignet als das Nullmodell.

Für die genaue Bedeutung der Koeffizienten des Nullmodells wird auf die Erläuterungen zur Tabelle 6 verwiesen. Die Nullmodelle aus Tabelle 6 und Tabelle 12 sind identisch.

Im zweiten Modell soll wiederum die Veränderung über die Zeit, diesmal jedoch jeweils getrennt für das erste und das zweite Projektjahr, untersucht werden. Anstelle der kontinuierlichen Zeitvariable werden die bereits im Abschnitt 6.5.3 erläuterten Dummy-Zeit-Variablen „J1“ und „J2“ modelliert. In diesem Modell werden noch keine Unterschiede zwischen der Interventions- und der Kontrollgruppe ermittelt. Der Achsenabschnitt  $\gamma_{000}$  (2.869\*\*\*) gibt stattdessen den Mittelwert aller Klassen zum ersten Messzeitpunkt an. Die zugehörigen Varianzkomponenten ( $r_{0ij} = 0.161^{***}$ ;  $u_{00j} = 0.015^{**}$ ) werden signifikant und die Varianz zwischen den Klassen ( $u_{00j} = 0.015^{**}$ ) fällt wider Erwarten größer aus als im Nullmodell. Wie bereits beim linearen Veränderungsmodell der schulischen Selbstwirksamkeitserwartung (siehe Tabelle 6) wird das Pseudo- $R^2$  der Varianzkomponente  $u_{00j}$  negativ (-0.875) und kann nicht interpretiert werden (Singer & Willett, 2003). Der

Koeffizient  $\gamma_{100}$  (-0.011) beschreibt die mittlere Abnahme der schulischen Selbstwirksamkeitserwartung aller Klassen vom ersten zum zweiten Messzeitpunkt. Diese wird genau wie die mittlere Zunahme der schulischen Selbstwirksamkeitserwartung aller Klassen im zweiten Projektjahr ( $\gamma_{200} = 0.041$ ) nicht signifikant. Die Schüler-Varianzkomponente  $r_{1ij}$  (0.051\*\*), die sich auf die Veränderung im ersten Projektjahr bezieht, wird signifikant. Wie in Abschnitt 6.5.3 beschrieben kann die Schüler-Varianzkomponente  $r_{2ij}$  für das zweite Projektjahr nicht modelliert werden, da das Modell sonst nicht iteriert. Dies gilt auch für die diskontinuierlichen Modelle der nachfolgenden Kriterien. Für die Frage nach der Wirksamkeit interessieren vor allem die Varianzkomponenten der Klassenebene, die sich auf die mittlere Veränderung im ersten ( $u_{10j} = 0.003$ ) und zweiten Projektjahr beziehen ( $u_{20j} = 0.006^*$ ). Diese fallen damit sehr klein aus und nur die Komponente für das zweite Projektjahr wird knapp signifikant.

Im dritten Modell, dem Evaluationsmodell, wird die jahresspezifische Veränderung der Interventionsgruppe im Vergleich zur Kontrollgruppe untersucht. Der Koeffizient  $\gamma_{000}$  (2.827\*\*\*) bildet den mittleren Wert aller Kontrollklassen zum ersten Messzeitpunkt ab und  $\gamma_{001}$  (0.084) die Differenz zwischen dem Mittelwert aller Interventionsklassen und dem Mittelwert aller Kontrollklassen zum ersten Messzeitpunkt. Somit erfasst der Parameter  $\gamma_{001}$  den Unterschied in den Ausgangswerten. Dieser ist nicht signifikant und die Zugehörigkeit zur Interventions- oder Kontrollgruppe trägt vom Veränderungs- zum Evaluationsmodell zu einer Varianzaufklärung bezüglich der Ausgangswerte von 6.7 % bei ( $u_{00j} = 0.014^{**}$ ; Pseudo-  $R^2$  von  $u_{00j} = 0.067$ ). Die durchschnittliche Veränderung aller Kontrollklassen über das erste Projektjahr erfasst  $\gamma_{100}$  (0.014), und der Koeffizient  $\gamma_{200}$  (0.022) bildet die durchschnittliche Veränderung aller Kontrollklassen über das zweite Projektjahr ab. In beiden Jahren verändern sich die Kontrollklassen nicht signifikant. Den Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen über das erste beziehungsweise das zweite Projektjahr beschreiben die Parameter  $\gamma_{101}$  (-0.050) und  $\gamma_{201}$  (0.041). Das negative Vorzeichen von  $\gamma_{101}$  weist zwar daraufhin, dass die schulische Selbstwirksamkeitserwartung in der Interventionsgruppe zum zweiten Messzeitpunkt um 0.05 Skaleneinheiten niedriger liegt als in der Kontrollgruppe ( $\gamma_{100} = 0.014$ ) und das positive Vorzeichen von  $\gamma_{201}$  verzeichnet für das zweite Jahr in der Interventionsgruppe eine etwas stärkere Zunahme in der Selbstwirksamkeit als in der Kontrollgruppe ( $\gamma_{100} = 0.022$ ). Zudem führt der Prädiktor ‚Gruppe‘ vom Veränderungs- zum Evaluationsmodell bezogen auf die Veränderung im ersten Pro-

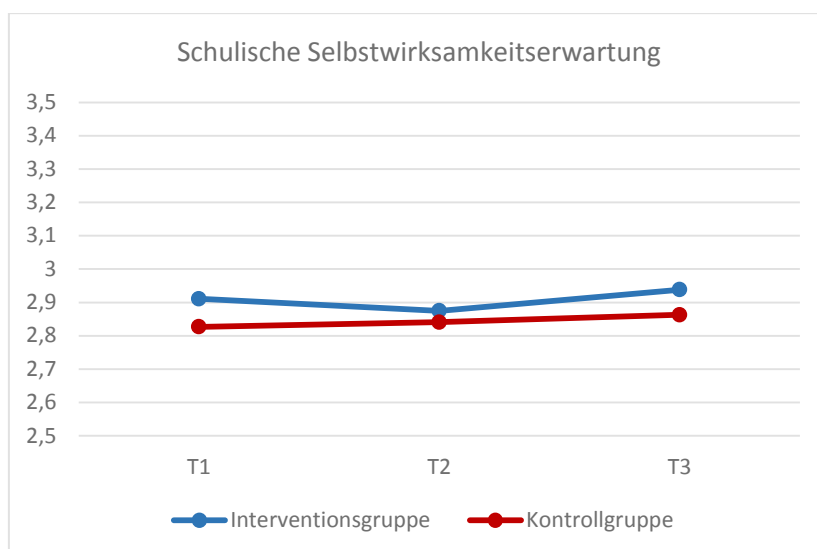
jektjahr zu einer Varianzaufklärung von 33.3 % und bezogen auf die Veränderung im zweiten Projektjahr zu einer Varianzaufklärung von 16.7 % ( $u_{10j} = 0.002$ ; Pseudo-  $R^2$  von  $u_{10j} = 0.333$ ;  $u_{20j} = 0.005^*$ ; Pseudo- $R^2$  von  $u_{20j} = 0.167$ ). Die Steigungskoeffizienten  $\gamma_{101}$  und  $\gamma_{201}$  sind jedoch nicht signifikant und verdeutlichen somit, dass sich Interventions- und Kontrollklassen weder im ersten noch im zweiten Projektjahr in ihrer mittleren Veränderung in der schulischen Selbstwirksamkeitserwartung bedeutsam unterscheiden. Da keine Prädiktoren auf der Schülerebene aufgenommen werden, verändern sich die Varianzkomponenten auf dieser Ebene  $r_{0ij}$  ( $0.161^{***}$ ) und  $r_{1ij}$  ( $0.051^{***}$ ) vom Veränderungs- zum Evaluationsmodell in der Regel nicht.

Zusammenfassend lässt sich festhalten: Bei ähnlichen Ausgangswerten zeigen sich hinsichtlich der Veränderung in der schulischen Selbstwirksamkeitserwartung in beiden Projektjahren keine bedeutsamen Unterschiede zwischen Interventions- und Kontrollgruppe. Abbildung 13 veranschaulicht dies grafisch: Interventionsgruppe (blaue Linie) und Kontrollgruppe (rote Linie) ähneln sich in ihrer jahresspezifischen Veränderung in der schulischen Selbstwirksamkeitserwartung. Nach den Kriterien zur Beurteilung der Modellgüte sind weder das Veränderungs- noch das Evaluationsmodell besser zur Vorhersage der Daten geeignet als das Nullmodell.

**Tabelle 12:** Diskontinuierliche, jahresspezifische Veränderung in der schulischen Selbstwirksamkeitserwartung

n L1 = 1355 n L2 = 566 n L3 = 22	Parameter	1. Nullmodell	2. Veränderungsmodell (J1&J2)	3. Evaluationsmodell (J1, J2 & Gruppe)
<b>Feste Effekte</b>				
Achsenabschnitt	$\gamma_{000}$	Koeffizient (SE) 2.879*** (0.03)	Koeffizient (SE) 2.869*** (0.04)	Koeffizient (SE) 2.827*** (0.06)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	-0.011 (0.03)	0.014 (0.04)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	0.041 (0.03)	0.022 (0.05)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	0.084 (0.07)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	-0.050 (0.05)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	0.041 (0.06)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.147	0.129	0.129
Level 2:	$r_{0ij}$	0.168***	0.161***	0.161***
	$r_{1ij}$	--	0.051***	0.051***
Level 3:	$u_{00j}$	0.008**	0.015**	0.014**
	$u_{10j}$	--	0.003	0.002
	$u_{20j}$	--	0.006*	0.005*
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	-0.875	0.067
	$u_{10j}$	--	--	0.333
	$u_{20j}$	--	--	0.167
<b>Modellfit</b>				
Devianz(q)		1989.443(4)	1975.646(13)	1973.804(16)
$\chi^2$ -Differenz			13.797	1.842
AIC		1997.443	2001.646	2005.804

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.025



**Abbildung 13: Diskontinuierliche, jahresspezifische Veränderung in der schulischen Selbstwirksamkeitserwartung (T1-T3 = Messzeitpunkt 1-3)**

### *Prüfungsängstlichkeit*

In Tabelle 13 ist die diskontinuierliche, jahresspezifische Veränderung in der Prüfungsängstlichkeit dargestellt. Wie bei den linearen Modellen weist das Evaluationsmodell eindeutig den besten Modellfit auf. Verglichen mit den entsprechenden Kennwerten des Null- (Devianz = 2761.368; AIC = 2769.368) und des Veränderungsmodells (Devianz = 2710.847; AIC = 2736.847) sind sowohl die Devianz (2698.673) als auch der AIC (2730.673) des Evaluationsmodells am kleinsten. Auch die Differenz zwischen dem Veränderungs- und dem Evaluationsmodell (12.174\*\*) wird im  $\chi^2$ -Test signifikant und spricht somit ebenfalls für die Güte des Evaluationsmodells. Nachfolgend sollen die Koeffizienten dieses besten Modells näher erläutert werden:

Die Ausgangswerte der Interventions- und Kontrollgruppe unterscheiden sich nicht signifikant ( $\gamma_{001} = -0.029$ ) und die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ trägt vom Veränderungs- zum Evaluationsmodell zu keiner Varianzaufklärung der Gruppenunterschiede in den Ausgangswerten bei (Pseudo- $R^2$  von  $u_{00j} = 0.000$ ). Wie auch schon bei den linearen Analysen (siehe Erläuterungen zur Tabelle 7) bleibt die Varianzkomponente  $u_{00j}$  (0.020\*\*) des diskontinuierlichen Evaluationsmodells verglichen mit dem Veränderungsmodell signifikant. Trotz der Aufnahme des Prädiktors ‚Gruppe‘ besteht demnach weiterhin eine substantielle Varianz zwischen den Klassen bezogen auf die Ausgangswerte, die jedoch nicht auf die Zugehörigkeit zur Interventions- oder Kontrollgruppe zurückgeführt werden kann. Der Parameter  $\gamma_{101}$  (-0.285\*\*) wird signifikant und beschreibt den Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren

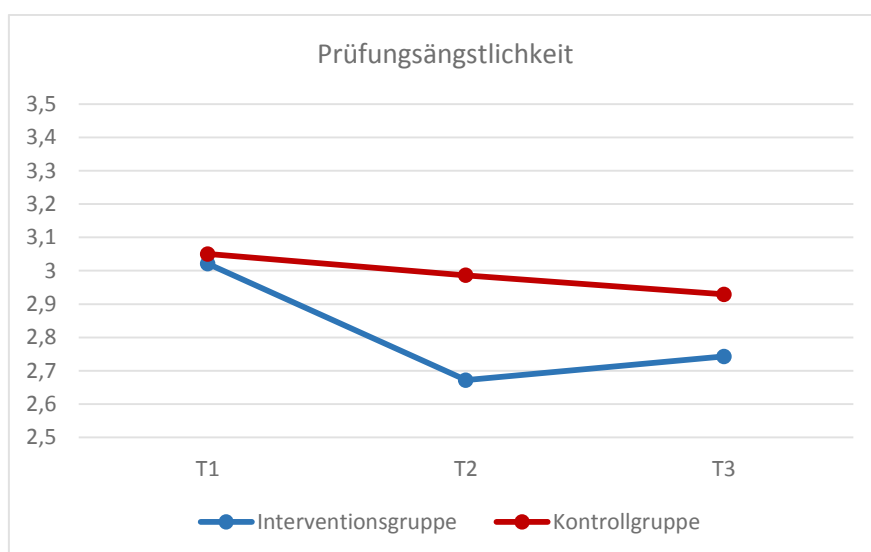
Veränderung aller Kontrollklassen über das erste Projektjahr. Die Prüfungsängstlichkeit liegt in der Interventionsgruppe zum zweiten Messzeitpunkt, also nach dem ersten Projektjahr, signifikant um 0.285 Skaleneinheiten niedriger als in der Kontrollgruppe. In dieser nimmt die Prüfungsängstlichkeit nach dem ersten Projektjahr zwar ebenfalls ab ( $\gamma_{100} = -0.064$ ), die Veränderung ist jedoch nicht signifikant. Der Prädiktor ‚Gruppe‘ führt vom Veränderungs- zum Evaluationsmodell zu einer Aufklärung der Varianz  $u_{10j}$  von 61.3 % (Pseudo- $R^2$  von  $u_{10j} = 0.613$ ) und die Varianzkomponente  $u_{10j}$  (0.012) ist im Evaluationsmodell nicht mehr signifikant. Parallel zu  $\gamma_{101}$  erfasst der Koeffizient  $\gamma_{201}$  (0.128) den Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen über das zweite Projektjahr. Im Gegensatz zur Kontrollgruppe ( $\gamma_{200} = -0.057$ ) nimmt die Prüfungsängstlichkeit in der Interventionsgruppe im zweiten Jahr wieder zu und der Prädiktor ‚Gruppe‘ klärt vom Veränderungs- zum Evaluationsmodell 33.3 % der Varianz bezogen auf die Veränderung im zweiten Jahr auf (Pseudo- $R^2$  von  $u_{20j} = 0.333$ ). Hierbei ist zu beachten, dass sich diese Varianzaufklärung auf die bereits im Veränderungsmodell sehr kleine und nicht signifikante Varianzkomponente  $u_{20j}$  (0.006) bezieht und der Effekt  $\gamma_{201}$  nicht signifikant ist. Interventions- und Kontrollklassen unterscheiden sich in ihrer Veränderung im zweiten Projektjahr folglich nicht bedeutsam. Wie bei den linearen Mehrebenenanalysen verändern sich die Varianzkomponenten der Schülerebene bis auf minimale durch die Schätzverfahren verursachte Abweichungen vom Veränderungs- zum Evaluationsmodell nicht, da im Zielmodell wiederum keine zusätzlichen Prädiktoren auf der Schülerebene aufgenommen werden. Dies gilt auch für die nachfolgenden diskontinuierlichen Modelle.

Zusammenfassend lässt sich festhalten: Bei nicht signifikant unterschiedlichen Ausgangswerten nimmt die Prüfungsängstlichkeit innerhalb des ersten Projektjahres in den Interventionsklassen signifikant stärker ab als in den Kontrollklassen. Anders als die linearen Modelle, die in der Interventionsgruppe eine Abnahme der Prüfungsängstlichkeit über beide Jahre beschreiben, weist das diskontinuierliche Evaluationsmodell im zweiten Jahr auf eine Zunahme der Prüfungsängstlichkeit in der Interventionsgruppe hin (siehe auch Abbildung 14: Interventionsgruppe = blaue Linie; Kontrollgruppe = rote Linie). Der Effekt ist jedoch nicht signifikant. Die Kriterien zur Beurteilung der Modellgüte beschreiben das Evaluationsmodell als das beste Modell.

**Tabelle 13:** Diskontinuierliche, jahresspezifische Veränderung in der Prüfungsängstlichkeit

n L1 = 1339 n L2 = 566 n L3 = 22	Parameter	1. Nullmodell	2. Veränderungs- modell (J1&J2)	3. <b>Evaluationsmodell (J1, J2 &amp; Gruppe)</b>
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.898*** (0.04)	3.037*** (0.05)	3.050*** (0.08)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	-0.203*** (0.05)	-0.064 (0.05)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	0.004 (0.04)	-0.057 (0.03)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.029 (0.09)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	-0.285** (0.09)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	0.128 (0.08)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.303	0.265	0.265
Level 2:	$r_{0ij}$	0.211***	0.194***	0.194***
	$r_{1ij}$	--	0.053	0.053
Level 3:	$u_{00j}$	0.023***	0.020**	0.020**
	$u_{10j}$	--	0.031**	0.012
	$u_{20j}$	--	0.006	0.004
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.130	0.000
	$u_{10j}$	--	--	0.613
	$u_{20j}$	--	--	0.333
<b>Modellfit</b>				
Devianz(q)		2761.368(4)	2710.847(13)	2698.673(16)
$\chi^2$ -Differenz		--	50.521***	12.174**
AIC		2769.368	2736.847	2730.673

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.043



**Abbildung 14:** Diskontinuierliche, jahresspezifische Veränderung in der Prüfungsängstlichkeit (T1-T3 = Messzeitpunkt 1-3)

### ***Schulbezogenes Stresserleben – wahrgenommene Herausforderung***

Tabelle 14 beschreibt die diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Herausforderung. Die Kriterien zur Beurteilung der Modellgüte weisen das Evaluationsmodell als das beste Modell aus. Im Vergleich zu den Gütekriterien des Null- (Devianz = 2777.068; AIC = 2785.068) und des Veränderungsmodells (Devianz = 2670.157; AIC = 2696.157) sind sowohl die Devianz (2654.087) als auch der AIC (2686.087) des Evaluationsmodells am geringsten. Zudem wird im  $\chi^2$ -Test zusätzlich die Differenz zwischen dem Veränderungs- und Evaluationsmodell (16.070\*\*\*) signifikant. Dies spricht ebenfalls für die Güte des Evaluationsmodells.

Der bereits bei den linearen Modellen der wahrgenommenen Herausforderung beschriebene signifikante Unterschied zwischen den Gruppen in den Ausgangswerten bleibt bestehen ( $\gamma_{001} = 0.323^{***}$ ). Verglichen mit der entsprechenden Varianzkomponente des Veränderungsmodells trägt die Aufnahme des Prädiktors ‚Gruppe‘ zu einer Aufklärung von 77.8 % bei (Pseudo- $R^2$  von  $u_{00j} = 0.778$ ) und die Varianz  $u_{00j}$  (0.006) ist im Evaluationsmodell nicht mehr signifikant. Kohärent zu den Ergebnissen der linearen Modelle fällt zu Beginn des Projekts die wahrgenommene Herausforderung in der Interventionsgruppe signifikant höher aus als in der Kontrollgruppe und der Unterschied kann durch die Zugehörigkeit zur Gruppe zu fast 80 % aufgeklärt werden. Der im linearen Evaluationsmodell noch signifikante Gruppeneffekt über die Zeit (siehe Erläuterungen der Tabelle 8) verliert sich bei der jahresspezifischen Analyse. Sowohl der Parameter  $\gamma_{101}$  (-0.173), der den Unterschied zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen für das erste Projektjahr beschreibt, als auch der parallel für das zweite Projektjahr zu interpretierende Parameter  $\gamma_{201}$  (-0.038) werden nicht signifikant. Beide Koeffizienten weisen jedoch ein negatives Vorzeichen auf und entsprechen somit verglichen mit der Kontrollgruppe dem bereits in den linearen Analysen verzeichneten stärker abnehmenden Trend der wahrgenommenen Herausforderung in der Interventionsgruppe. Für die Kontrollschüler ist im ersten Projektjahr eine weniger starke Reduktion der Herausforderung zu vermerken ( $\gamma_{100} = -0.191^{**}$ ).<sup>50</sup> Im zweiten Jahr nimmt die Herausforderung in der Kontrollgruppe – anders als in der Interventionsgruppe – sogar minimal zu

---

<sup>50</sup> Der im Vergleich zum Parameter  $\gamma_{101}$  (-0.173) größere absolute Betrag des Parameters  $\gamma_{100}$  (-0.191) könnte zu der falschen Annahme verleiten,  $\gamma_{100}$  verzeichne eine stärkere Reduktion in der wahrgenommenen Herausforderung für die Kontrollgruppe als  $\gamma_{101}$ . Deshalb sei an dieser Stelle daran erinnert, dass die Parameter  $\gamma_{101}$  und  $\gamma_{201}$  nicht die Veränderung der Interventionsgruppe im jeweiligen Projektjahr erfassen, sondern den Unterschied zwischen der mittleren Veränderung in der Interventionsgruppe und der mittleren Veränderung in der Kontrollgruppe abbilden.



( $\gamma_{200} = 0.017$ ). Diese Veränderung ist jedoch nicht signifikant. Der Prädiktor ‚Gruppe‘ klärt vom Veränderungs- zum Evaluationsmodell 33 % der Varianz bezogen auf die Veränderung im ersten Jahr auf (Pseudo- $R^2$  von  $u_{10j} = 0.333$ ). Das Pseudo- $R^2$  von  $u_{20j}$  (-2) für das zweite Jahr wird negativ und kann deshalb nicht interpretiert werden (Singer & Willett, 2003).

Zusammenfassend nimmt die wahrgenommene Herausforderung in beiden Projektjahren bei den Interventionsschülern stärker ab als bei den Kontrollschülern. Die Unterschiede sind jedoch nicht signifikant und müssen in der Diskussion auch vor dem Hintergrund der signifikant unterschiedlichen Ausgangswerte interpretiert werden. Im Hinblick auf den Modellfit weist das Evaluationsmodell die beste Vorhersagefähigkeit auf.

**Tabelle 14:** Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Herausforderung

n L1 = 1352 n L2 = 566 n L3 = 22		1.	2.	3.
	Parameter	Nullmodell	Veränderungsmodell (J1&J2)	Evaluationsmodell (J1, J2 & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.382*** (0.04)	2.576*** (0.05)	2.414*** (0.05)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	-0.280*** (0.05)	-0.191** (0.05)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	0.001 (0.03)	0.017 (0.04)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	0.323*** (0.07)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	-0.173 (0.09)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	-0.038 (0.06)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.298	0.206	0.205
Level 2:	$r_{0ij}$	0.215***	0.277***	0.274***
	$r_{1ij}$	--	0.205***	0.203***
Level 3:	$u_{00j}$	0.027***	0.027***	0.006
	$u_{10j}$	--	0.015*	0.010
	$u_{20j}$	--	0.001	0.003
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.000	0.778
	$u_{10j}$	--	--	0.333
	$u_{20j}$	--	--	-2
<b>Modellfit</b>				
Devianz(q)		2777.068(4)	2670.157(13)	2654.087 (16)
$\chi^2$ -Differenz		--	106.911***	16.070***
AIC		2785.068	2696.157	2686.087

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; ICC für Level 3: 0.05

### ***Schulbezogenes Stresserleben – wahrgenommene Bedrohung***

In Tabelle 15 ist die diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Bedrohung dargestellt. Nach den Kriterien zur Beurteilung der Modellgüte ist das Evaluationsmodell am besten zur Vorhersage der Daten geeignet. Sowohl die Devianz (2671.300) als auch der AIC (2703.300) des Evaluationsmodells fallen verglichen mit den entsprechenden Koeffizienten des Null- (Devianz = 2738.082; AIC = 2746.082) und des Veränderungsmodells (Devianz = 2680.694; AIC = 2706.694) am geringsten aus. Auch der  $\chi^2$ -Test weist auf die Güte des Evaluationsmodells hin, denn die Differenz zwischen dem Veränderungs- und Evaluationsmodell (9.394\*) wird signifikant.

Wie bereits im linearen Evaluationsmodell (siehe Erläuterungen zur Tabelle 9) unterscheiden sich die Ausgangswerte zwischen Interventions- und Kontrollgruppe nicht signifikant ( $\gamma_{001} = -0.021$ ). Anders aber als im linearen Evaluationsmodell führt die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ vom diskontinuierlichen Veränderungs- zum Evaluationsmodell zu keiner Varianzaufklärung der Gruppenunterschiede in den Ausgangswerten (Pseudo- $R^2$  von  $u_{00j} = 0.000$ ) und die Varianzkomponente  $u_{00j}$  (0.009\*) des Evaluationsmodells bleibt verglichen mit dem Veränderungsmodell unverändert und signifikant. Folglich besteht weiterhin eine noch zu erklärende substantielle Variation zwischen den Klassen bezogen auf die Ausgangswerte, die nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe erklärt werden kann. Der Koeffizient  $\gamma_{101}$  (-0.211\*) ist negativ und wird signifikant. Verglichen mit den Kontrollschülern reduziert sich die wahrgenommene Bedrohung demnach bei den Interventionsschülern im ersten Projektjahr signifikant. Im Gegensatz dazu nimmt die wahrgenommene Bedrohung bei den Kontrollschülern im ersten Jahr sogar tendenziell zu ( $\gamma_{100} = 0.066$ ) – der Effekt ist jedoch nicht signifikant. Die Zugehörigkeit zur Interventions- oder Kontrollgruppe erklärt verglichen mit dem Veränderungsmodell 47.8 % der Varianz bezogen auf die Veränderung im ersten Jahr auf (Pseudo- $R^2$  von  $u_{10j} = 0.478$ ), wobei die Varianzkomponente  $u_{10j}$  (0.012\*) im Evaluationsmodell noch signifikant bleibt. Dies bedeutet, bezogen auf die Veränderung im ersten Jahr existiert weiterhin eine substantielle Variation zwischen den Klassen, die nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe aufgeklärt werden kann. Im zweiten Jahr nimmt die wahrgenommene Bedrohung sowohl in der Kontroll- als auch in der Interventionsgruppe ab, in Letzterer etwas stärker als in Ersterer ( $\gamma_{200} = -0.071$ ;  $\gamma_{201} = -0.047$ ).<sup>51</sup>

---

<sup>51</sup> Auch hier könnte der im Vergleich zum Parameter  $\gamma_{200}$  (-0.071) kleinere Betrag des Parameters  $\gamma_{201}$  (-0.047) zu der falschen Annahme verleiten,  $\gamma_{201}$  verzeichne eine geringere Reduktion in der wahrgenommenen Bedrohung als  $\gamma_{200}$ .

Dieser Gruppenunterschied in der Veränderung ist jedoch nicht signifikant, die zugehörige Varianzkomponente  $u_{20j}$  ( $0.010^*$ ) ist noch immer signifikant und vom Veränderungs- zum Evaluationsmodell trägt der Prädiktor ‚Gruppe‘ auch nur zu einer Aufklärung der entsprechenden Varianz von 9.1 % bei (Pseudo- $R^2$  von  $u_{20j} = 0.091$ ). Auch bezogen auf die Veränderung im zweiten Jahr besteht also weiterhin eine substantielle Variation zwischen den Klassen, die nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe aufgeklärt werden kann. Die in den linearen Modellen verzeichnete Reduktion des Bedrohungslebens in der Interventionsgruppe über beide Projektjahre hinweg ist folglich vor allem auf das erste Projektjahr zurückzuführen.

Zusammenfassend lässt sich festhalten: Bei nicht signifikant unterschiedlichen Ausgangswerten nimmt die wahrgenommene Bedrohung in den Interventionsklassen im ersten Projektjahr ab und in den Kontrollklassen zu. Der Unterschied ist signifikant. Im zweiten Projektjahr zeigen sich keine weiteren Gruppenunterschiede zwischen Interventions- und Kontrollgruppe. Nach den Kriterien zur Beurteilung der Modellgüte weist das Evaluationsmodell den besten Modellfit auf.

---

Deshalb sei an dieser Stelle erneut daran erinnert, dass die Parameter  $\gamma_{101}$  und  $\gamma_{201}$  nicht die Veränderung der Interventionsgruppe im jeweiligen Projektjahr erfassen, sondern den Unterschied zwischen der mittleren Veränderung in der Interventionsgruppe und der mittleren Veränderung in der Kontrollgruppe.

**Tabelle 15:** Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Bedrohung

n L1 = 1353 n L2 = 566 n L3 = 22		1.	2.	3.
Parameter		Nullmodell	Veränderungsmodell (J1&J2)	Evaluationsmodell (J1, J2 & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.157*** (0.04)	2.213*** (0.04)	2.222*** (0.05)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	-0.037 (0.05)	0.066 (0.05)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	-0.092* (0.04)	-0.071 (0.06)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.021 (0.07)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	-0.211* (0.09)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	-0.047 (0.08)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.296	0.254	0.254
Level 2:	$r_{0ij}$	0.195***	0.122***	0.122***
	$r_{1ij}$	--	0.084**	0.084**
Level 3:	$u_{00j}$	0.022***	0.009*	0.009*
	$u_{10j}$	--	0.023*	0.012*
	$u_{20j}$	--	0.011*	0.010*
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.591	0.000
	$u_{10j}$	--	--	0.478
	$u_{20j}$	--	--	0.091
<b>Modellfit</b>				
Devianz(q)		2738.082(4)	2680.694(13)	2671.300(16)
$\chi^2$ -Differenz		--	57.388***	9.394*
AIC		2746.082	2706.694	2703.300

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.043

### **Schulbezogenes Stresserleben – wahrgenommener Verlust**

Tabelle 16 gibt Auskunft über die diskontinuierliche, jahresspezifische Veränderung im wahrgenommenen Verlust. Veränderungs- und Evaluationsmodell unterscheiden sich kaum hinsichtlich ihrer Modellgüte. Zwar ist die Devianz des Evaluationsmodells (2901.344) kleiner als die des Veränderungsmodells (2907.764). Die Differenz der beiden Devianzen (6.420) wird jedoch nicht signifikant und in ihrem AIC unterscheiden sich beide Modelle nur in den Nachkommastellen (2933.764; 2933.344). Verglichen mit den Kennwerten des Nullmodells (Devianz = 2947.758; AIC = 2955.758) sind Veränderungs- und Evaluationsmodell besser zur Vorhersage geeignet. Auch die Differenz zwischen Null- und Veränderungsmodell wird signifikant (39.994\*\*\*).

Wie auch schon im linearen Evaluationsmodell des wahrgenommenen Verlusts (siehe Tabelle 10) unterscheiden sich die Ausgangswerte zwischen der Interventions- und der Kon-

trollgruppe auch hier im diskontinuierlichen Evaluationsmodell ( $\gamma_{001} = -0.140^*$ ) signifikant und die Aufnahme des Level-3-Prädiktors ‚Gruppe‘ führt vom Veränderungs- zum Evaluationsmodell zu einer Varianzaufklärung von 75 % der Gruppenunterschiede in den Ausgangswerten (Pseudo- $R^2$  von  $u_{00j} = 0.75$ ). Die Interventionsgruppe nimmt zu Beginn des Projekts signifikant weniger Verlust wahr als die Kontrollgruppe. Im Hinblick auf die Frage nach den jahresspezifischen Gruppenunterschieden in der Veränderung zeigt sich für das erste Projektjahr ein Anstieg des wahrgenommenen Verlusts in beiden Gruppen. Dieser ist jedoch in der Interventionsgruppe etwas weniger stark als in der Kontrollgruppe ( $\gamma_{100} = 0.214^{**}$ ;  $\gamma_{101} = -0.039$ ). Im zweiten Jahr reduziert sich das Verlusterleben in beiden Gruppen, in der Interventionsgruppe jedoch weniger stark als in der Kontrollgruppe ( $\gamma_{200} = -0.072$ ;  $\gamma_{201} = 0.040$ ).<sup>52</sup> Die Gruppenunterschiede in der Veränderung sind jedoch nicht bedeutsam und für das zweite Jahr werden beide Koeffizienten nicht signifikant. Im zweiten Jahr ist folglich über alle Klassen hinweg keine bedeutsame Veränderung zu verzeichnen. Verglichen mit dem Veränderungsmodell leistet der Prädiktor ‚Gruppe‘ keinen Beitrag zur Varianzaufklärung bezüglich der Veränderung im ersten Jahr (Pseudo- $R^2$  von  $u_{10j} = 0.000$ ) und der Veränderung im zweiten Jahr (Pseudo- $R^2$  von  $u_{20j} = 0.000$ ).

Zusammenfassend lässt sich festhalten, dass sich Interventions- und Kontrollklassen hinsichtlich ihres Verlusterlebens in beiden Projektjahren nicht signifikant unterscheiden. Bei der späteren Diskussion des Ergebnisses sind die signifikant unterschiedlichen Ausgangswerte zu beachten. Veränderungs- und Evaluationsmodell scheinen sich in ihrer Modellgüte kaum zu unterscheiden.

---

<sup>52</sup> Wieder sei an dieser Stelle zur Vermeidung von Irritationen daran erinnert, dass die Parameter  $\gamma_{101}$  und  $\gamma_{201}$  nicht die Veränderung der Interventionsgruppe im jeweiligen Projektjahr erfassen, sondern den Unterschied zwischen der mittleren Veränderung in der Interventionsgruppe und der mittleren Veränderung in der Kontrollgruppe abbilden.

**Tabelle 16:** Diskontinuierliche, jahresspezifische Veränderung im wahrgenommenen Verlust

n L1 = 1353 n L2 = 565 n L3 = 22		1.	2.	3.
Parameter		Nullmodell	Veränderungsmodell (J1&J2)	Evaluationsmodell (J1, J2 & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	1.782*** (0.04)	1.662*** (0.03)	1.734*** (0.04)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	0.197*** (0.05)	0.214** (0.06)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	-0.052 (0.04)	-0.072 (0.06)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.140* (0.05)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	-0.039 (0.09)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	0.040 (0.08)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.359	0.310	0.310
Level 2:	$r_{0ij}$	0.208***	0.193***	0.189***
	$r_{1ij}$	--	0.114***	0.114***
Level 3:	$u_{00j}$	0.013**	0.002	0.0005
	$u_{10j}$	--	0.010	0.0010
	$u_{20j}$	--	0.004	0.004
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.846	0.750
	$u_{10j}$	--	--	0.000
	$u_{20j}$	--	--	0.000
<b>Modellfit</b>				
Devianz(q)		2947.758(4)	2907.764(13)	2901.344(16)
$\chi^2$ -Differenz		--	39.994***	6.420
AIC		2955.758	2933.764	2933.344

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.022

### Wahrgenommene Binnendifferenzierung

In Tabelle 17 wird die diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Binnendifferenzierung im Unterricht dargestellt. Anders als bei den linearen Modellen bezogen auf die wahrgenommene Binnendifferenzierung (siehe Tabelle 11), wo sich Veränderungs- und Evaluationsmodell hinsichtlich ihrer Modellgüte kaum unterscheiden, weist bei der diskontinuierlichen Modellierung das Evaluationsmodell den besten Modellfit auf. Sowohl die Devianz (2540.312) als auch der AIC (2572.312) des Evaluationsmodells sind kleiner als die entsprechenden Kennwerte des Null- (Devianz = 2622.281; AIC = 2630.281) und des Veränderungsmodells (Devianz = 2554.138; AIC = 2580.138). Auch die Differenz der Devianzen des Veränderungs- und des Evaluationsmodells (13.826\*\*) ist signifikant und weist somit ebenfalls auf die Güte des Evaluationsmodells hin.

Die Ausgangswerte zwischen Interventions- und Kontrollgruppe unterscheiden sich nicht signifikant ( $\gamma_{001} = -0.037$ ) und die Varianzaufklärung vom Veränderungs- zum Evaluationsmodell bezogen auf die Ausgangswerte kann aufgrund eines negativen Pseudo- $R^2$  von  $u_{00j}$  (-0.417) nicht ermittelt werden. Verglichen mit dem Veränderungsmodell bleibt die Varianzkomponente  $u_{00j}$  (0.017\*) im Evaluationsmodell signifikant. Trotz der Aufnahme des Prädiktors ‚Gruppe‘ besteht folglich weiterhin eine substantielle Varianz bezogen auf die Ausgangswerte, die nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe erklärt werden kann. Die Parameter  $\gamma_{101}$  (0.344\*\*\*) und  $\gamma_{201}$  (-0.126) beschreiben den jahresspezifischen Unterschied in der wahrgenommenen Binnendifferenzierung zwischen der mittleren Veränderung aller Interventionsklassen und der mittleren Veränderung aller Kontrollklassen. Vom ersten zum zweiten Messzeitpunkt nehmen die Interventionschüler verglichen mit den Kontrollschülern durchschnittlich signifikant mehr Binnendifferenzierung im Unterricht wahr. Die Kontrollschüler geben im ersten Projektjahr sogar eine Reduktion der Binnendifferenzierung an ( $\gamma_{100} = -0.062$ ). Der Prädiktor ‚Gruppe‘ trägt somit vom Veränderungs- zum Evaluationsmodell zu einer Aufklärung der Varianz bezogen auf die Veränderung im ersten Projektjahr von 53.8 % bei (Pseudo- $R^2$  von  $u_{10j} = 0.538$ ) und die Varianzkomponente  $u_{10j}$  (0.012) ist im Evaluationsmodell nicht mehr signifikant. Vom zweiten zum dritten Messzeitpunkt nimmt die wahrgenommene Binnendifferenzierung im Unterricht der Interventionsklassen im Vergleich zu den Kontrollklassen wieder ab. Im Gegensatz dazu berichten die Kontrollschüler im zweiten Jahr eine Zunahme der Binnendifferenzierung ( $\gamma_{200} = 0.103$ ). Diese ist jedoch nicht signifikant und auch der Gruppenunterschied in der Veränderung des zweiten Projektjahres ist nicht bedeutsam. Die entsprechende Varianzaufklärung beträgt vom Veränderungs- zum Evaluationsmodell 6.3 % (Pseudo- $R^2$  von  $u_{20j} = 0.063$ ) und die Varianzkomponente  $u_{20j}$  (0.015\*) ist im Evaluationsmodell noch immer signifikant. Bezüglich der Veränderung vom zweiten zum dritten Projektjahr besteht somit weiterhin eine substantielle Variation zwischen den Klassen, die jedoch nicht durch die Zugehörigkeit zur Interventions- oder Kontrollgruppe aufgeklärt werden kann.

Zusammenfassend zeigt sich in der Interventionsgruppe verglichen mit der Kontrollgruppe bei nicht signifikant unterschiedlichen Ausgangswerten ein bedeutsamer Anstieg der wahrgenommenen Binnendifferenzierung im ersten Projektjahr. Im zweiten Jahr nimmt die Binnendifferenzierung in der Interventionsgruppe – anders als durch die lineare Modellierung (siehe Tabelle 11) zu erwarten gewesen wäre – wieder ab. Dieser Gruppenunterschied

in der Veränderung bezogen auf das zweite Projektjahr ist jedoch nicht signifikant. Als bestes Modell erweist sich das Evaluationsmodell.

**Tabelle 17:** Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Binnendifferenzierung

n L1 = 1341 n L2 = 566 n L3 = 22		1.	2.	3.
Parameter		Nullmodell	Veränderungsmodell (J1&J2)	Evaluationsmodell (J1, J2 & Gruppe)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{000}$	2.136*** (0.04)	2.052*** (0.04)	2.073*** (0.06)
Veränderung vom 1. zum 2. Projektjahr	$\gamma_{100}$	--	0.105 (0.05)	-0.062 (0.04)
Veränderung vom 2. zum 3. Projektjahr	$\gamma_{200}$	--	0.043 (0.04)	0.103 (0.06)
Gruppenunterschied im Ausgangswert	$\gamma_{001}$	--	--	-0.037 (0.09)
Gruppenunterschied in der Veränderung vom 1. zum 2. Projektjahr	$\gamma_{101}$	--	--	0.334*** (0.08)
Gruppenunterschied in der Veränderung vom 2. zum 3. Projektjahr	$\gamma_{201}$	--	--	-0.126 (0.08)
<b>Zufallseffekte (Varianzkomponenten)</b>				
Level 1:	$e_{tij}$	0.331	0.237	0.235
Level 2:	$r_{0ij}$	0.088***	0.248***	0.244***
	$r_{1ij}$	--	0.233***	0.233***
Level 3:	$u_{00j}$	0.025***	0.012*	0.017*
	$u_{10j}$	--	0.026*	0.012
	$u_{20j}$	--	0.016*	0.015*
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>				
Level 3:	$u_{00j}$	--	0.520	-0.417
	$u_{10j}$	--	--	0.538
	$u_{20j}$	--	--	0.063
<b>Modellfit</b>				
Devianz(q)		2622.281(4)	2554.138(13)	2540.312(16)
$\chi^2$ -Differenz		--	68.143***	13.826**
AIC		2630.281	2580.138	2572.312

Anmerkungen: n = Stichprobengröße; L1-3 = Level 1-3; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 3: 0.056

### 7.1.3 Zusammenfassung der Ergebnisse der Wirksamkeitsevaluation

Zusammenfassend findet sich im Rahmen der linearen Wachstumskurvenmodelle ein erwartungskonformer signifikanter Effekt in der wahrgenommenen Bedrohung: Diese reduziert sich im Laufe der beiden Projektjahre in der Interventionsgruppe stärker als in der Kontrollgruppe. Entgegen der Erwartung nimmt das Herausforderungserleben in den Interventionsklassen über die Zeit signifikant stärker ab als in den Kontrollklassen. Die Betrachtung der jahresspezifischen Veränderung ergibt ein differenzierteres Bild: Zusätzlich zur wahrgenommenen Bedrohung ergeben sich zwei weitere signifikante erwartungskonforme Effekte in der Prüfungsängstlichkeit und der erlebten Binnendifferenzierung. Am



Ende des ersten Projektjahres fällt Erstere bei den Interventionsschülern signifikant geringer, Letztere bedeutsam höher aus als bei den Kontrollschülern. Auch der Effekt in der wahrgenommenen Bedrohung zeigt sich im Rahmen der diskontinuierlichen Modellierung und jahresspezifischen Überprüfung nur für das erste Jahr. Der erwartungswidrige Effekt in der wahrgenommenen Herausforderung verliert sich bei der jahresspezifischen Betrachtung. Eine ausführliche Interpretation dieser Ergebnisse erfolgt in der Diskussion (Kapitel 8). Nachfolgend wird zunächst der Einfluss der wahrgenommenen Implementation auf die Zielkriterien untersucht.

## **7.2 Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation**

Bevor die Ergebnisse der kovarianzanalytischen Mehrebenenmodelle zur Analyse der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler erläutert werden, soll zunächst die Programmumsetzung deskriptiv beschrieben werden. Auf der Basis der Beteiligung der Lehrkräfte an den Workshops können indirekt Erkenntnisse zur Umsetzungshäufigkeit der Workshopinhalte in den Fachunterricht abgeleitet werden. Auch die Anzahl in Anspruch genommener Beratungssitzungen zur Umsetzung der Workshopinhalte in den Schulen gibt Aufschluss über eine mögliche Innovationsbereitschaft der Lehrkräfte. Zudem wird die wahrgenommene Implementation seitens der Schüler deskriptiv dargestellt. Diese liefert ebenfalls Informationen über die Umsetzung der verschiedenen Förderstrategien aus der Perspektive der Schüler.

### **7.2.1 Deskriptive Beschreibung der Implementation**

Die Beteiligung der Lehrkräfte an den Workshops fällt im zweiten Projektjahr geringer aus als im ersten (siehe Tabelle 18). Über alle Schulen hinweg beteiligten sich am ersten Workshop zum Thema Klassenklima 39 Lehrkräfte. Am zweiten und dritten Workshop zu den Themen Transparenz und Lern- und Leistungsraum nahmen bereits etwas weniger, jeweils 24 und 25, Lehrkräfte teil. Im zweiten Jahr sank die Beteiligung noch weiter auf jeweils 17, 18 und 19 Lehrkräfte pro Workshop (Tabelle 18).<sup>53</sup>

---

<sup>53</sup> Bei der Bewertung der Beteiligungsdaten ist zu beachten, dass die Schulleitung der Projektschule 3 zu Beginn des FOSIS2-Projekts das gesamte Kollegium und nicht nur die Zielgruppe, die in den siebten Klassen unterrichtenden Lehrkräfte, aufgefordert hatte, am ersten Workshop des FOSIS2-Projekts teilzunehmen. Erst ab dem zweiten Workshop nahmen an dieser Schule nur die Lehrkräfte der Schule an dem Projekt teil, die in den siebten Klassen unterrichteten. So erklärt sich der Rückgang der Beteiligung um circa 50 % an der Schule 3

**Tabelle 18:** Beteiligung der Lehrkräfte an den Workshops

Workshops	1. Projektjahr			2. Projektjahr		
	KK	TP	LLR	SB	KL	PHM
Schule 1	14	10	7	4	4	4
Schule 2	11	7	10	6	8	8
Schule 3	14	7	8	7	6	7
<b>gesamt</b>	39	24	25	17	18	19

*Anmerkungen:* KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum; SB = Selbstbestimmung; KL = Kooperatives Lernen; PHM = Selbstregulation (proaktives Handlungsregulationsmodell)

Auch die Inanspruchnahme von modulspezifischen Beratungen (siehe Tabelle 19) durch die wissenschaftliche Begleitung ist im zweiten Projektjahr geringer als im ersten. Im ersten Projektjahr zeigt sich über alle Schulen hinweg für das erste Modul Klassenklima die häufigste Inanspruchnahme ( $n = 10$ ). Zum Modul Transparenz, das mit der Maßnahme Transparenzpapier eine für die Lehrkräfte sehr leicht in den Unterricht zu implementierende Förderstrategie beinhaltet, fand nur eine Beratungssitzung in der Schule 2 statt. Für die Einführung des sogenannten Lernraumes, eine Maßnahme, die stärker in die Unterrichtsroutinen eingreift als das Transparenzpapier, wurden wieder etwas mehr ( $n = 4$ ) Beratungssitzungen in Anspruch genommen. Im Vergleich zu den im ersten Jahr insgesamt 15 stattfindenden Beratungssitzungen wurden im zweiten Jahr deutlich weniger ( $n = 6$ ) Beratungssitzungen umgesetzt: Für die eher einfache Implementation der Inhalte des Moduls Selbstbestimmung (Wahlmöglichkeiten im Unterricht) wurde eine Beratungssitzung in der Schule 3 realisiert. Obwohl insbesondere die Module zu den Themen Kooperatives Lernen und Selbstregulation als die komplexesten Module bewertet werden können, bei deren Umsetzung die meisten Schwierigkeiten zu erwarten waren, wurden für die Module nur jeweils 3 und 2 Beratungssitzungen in Anspruch genommen (Tabelle 19).

**Tabelle 19:** Anzahl an modulspezifischen Beratungssitzungen im Anschluss an die Workshops

Workshops	1. Projektjahr			2. Projektjahr		
	KK	TP	LLR	SB	KL	PHM
Schule 1	2	--	1	--	--	--
Schule 2	3	1	1	--	1	1
Schule 3	5	--	2	1	2	1
<b>gesamt</b>	10	1	4	1	3	2
<b>gesamt pro Jahr</b>	15			6		

*Anmerkungen:* KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum; SB = Selbstbestimmung; KL = Kooperatives Lernen; PHM = Selbstregulation (proaktives Handlungsregulationsmodell)

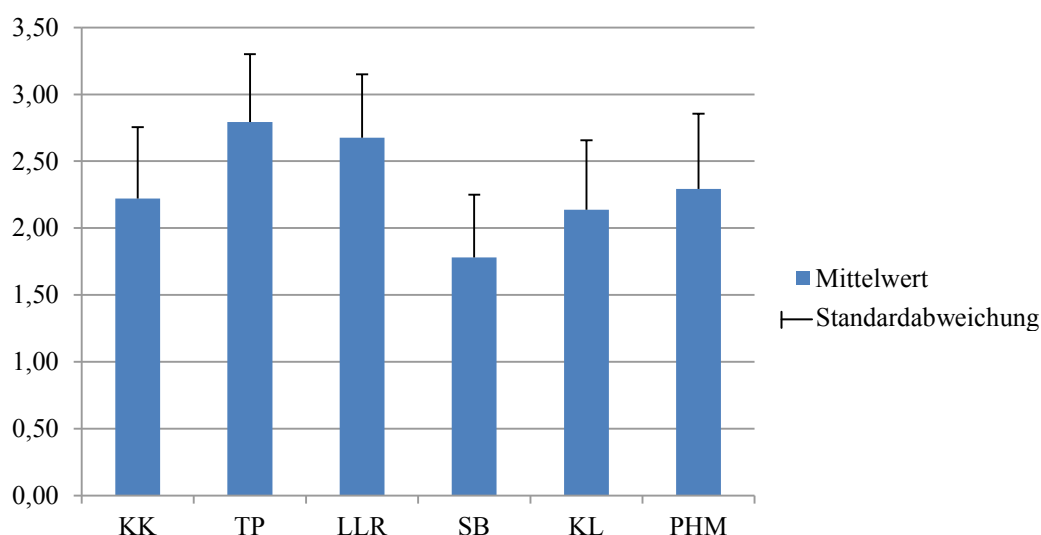
Bei Betrachtung der deskriptiven Statistik der von den Schülern wahrgenommenen modulspezifischen Implementation zeigt sich ein ähnliches Bild (siehe Tabelle 20 und Abbildung 15). Die wahrgenommene Implementation fällt im ersten Projektjahr höher aus als im zweiten. Am höchsten ist die Wahrnehmung der Umsetzung der Förderstrategie Transparenzpapier ( $M = 2.79$ ), gefolgt von der Umsetzung eines Lernraumes im Unterricht ( $M = 2.68$ ). Die wahrgenommene Implementation der Inhalte der Module Klassenklima ( $M = 2.22$ ) und Selbstregulation ( $M = 2.29$ ) ist fast gleich hoch. Noch weniger als diese Inhalte werden die Maßnahmen der Module Kooperatives Lernen ( $M = 2.14$ ) und Selbstbestimmung ( $M = 1.78$ ) wahrgenommen.

Insgesamt ist im zweiten Projektjahr folglich von einer geringeren Umsetzung der Förderstrategien auszugehen als im ersten. Eine ausführliche Diskussion der deskriptiven Beschreibung der Implementation findet sich in Kapitel 8.2.1.

**Tabelle 20:** Wahrgenommene modulspezifische Implementation seitens der Schüler

	n	M	SD	Range
Klassenklima	207	2.22	0.53	1.0 - 3.5
Transparenz	212	2.79	0.51	1.32 - 3.95
Lern- und Leistungsraum	209	2.68	0.48	1.2 - 3.92
Selbstbestimmung	205	1.78	0.47	1.0 - 3.38
Kooperatives Lernen	204	2.14	0.51	1.0 - 3.10
Selbstregulation	213	2.29	0.56	1.0 - 3.81

Anmerkungen: n = Stichprobengröße; M = Mittelwert; SD = Standardabweichung



**Abbildung 15:** Wahrgenommene modulspezifische Implementation seitens der Schüler

Anmerkungen: KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum; SB = Selbstbestimmung; KL = Kooperatives Lernen; PHM = Selbstregulation (proaktives Handlungsregulationsmodell)

### 7.2.2 Kovarianzanalytische Veränderungsmodelle

Zur Analyse der Frage nach der Wirksamkeit in Abhängigkeit von der Programmumsetzung werden der modulspezifische Einfluss der wahrgenommenen Implementation seitens der Schüler für das erste Projektjahr untersucht und die unter 6.5.4 erläuterten modulspezifischen Implementationsprädiktoren (,ImpS.kk‘, ,ImpS.tp‘ und ,ImpS.llr‘) auf der Schülerebene sukzessiv modelliert. Die Beschränkung der Analysen auf das erste Projektjahr und auch die Wahl der Implementationsindikatoren wurden bereits zu Beginn der Modellbildung in Kapitel 6.5.4 begründet. Im Unterschied zu den linearen und diskontinuierlichen Veränderungsmodellen werden bei den Berechnungen zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation folglich nur zwei Messzeitpunkte berücksichtigt, die Prä- und die Zwischen-Befragung. Damit entfällt die Ebene der Zeit und es wird ein kovarianzanalytisches Zwei-Ebenen-Modell durchgeführt: mit der Ebene der Schüler als Level 1 und der Ebene der Klassen als Level 2. Das Implementationsgeschehen konnte nur in den Interventionsklassen erfragt werden, da in den Kontrollklassen keine Umsetzung der Maßnahmen erfolgte. Für die kovarianzanalytischen Modelle besteht die zugrunde liegende Stichprobe deshalb nur aus Interventionsschülern. Anders als bei den linearen und diskontinuierlichen Modellen ist bei den Modellen zur Prüfung des Implementationseinflusses außerdem zu beachten, dass hier kein Prädiktor auf der Klassenebene, stattdessen aber teilweise mehrere modulspezifische Prädiktoren auf der Schülerebene aufgenommen werden. Die Varianz zwischen den Klassen ist für die Frage nach dem Einfluss der wahrgenommenen Implementation seitens der Schüler somit von geringem Interesse und wird deshalb mittels der ,Group-Mean-Zentrierung‘ nivelliert (Eid et al., 2013; siehe auch Anhang C2). Anders als bei den linearen und diskontinuierlichen Modellen kann es bei den kovarianzanalytischen Modellen mehrere miteinander konkurrierende Zielmodelle geben. Denn durch die Möglichkeit, Modelle verschiedener Kombinationen von Implementationsprädiktoren zu bilden, existieren mehrere Zielmodelle, die zur Beantwortung der Forschungsfrage herangezogen werden können (siehe auch die ausführliche Beschreibung der Modellbildung unter 6.5.4). Zur Ermittlung des besten Modells kann der AIC für den Vergleich aller Modelle, auch miteinander konkurrierender Zielmodelle, herangezogen werden. Der  $\chi^2$ -Test hingegen darf nur für den Vergleich ineinander geschachtelter Modelle verwendet werden (siehe Anhang C2). Auch das Pseudo- $R^2$  wird jeweils aus den in sich geschachtelten Modellen ermittelt: bei Modellen mit einem Implementationsprädiktor im Vergleich zum autoregressiven Modell und bei Modellen mit

zwei Implementationsprädiktoren im Vergleich zum Modell mit nur einem. Bei der tabellarischen Darstellung sind die geschachtelten Modelle mit einer jeweils höheren Nummer gekennzeichnet (z. B. 1., 2., 3.). Nicht geschachtelte, miteinander konkurrierende Zielmodelle sind mit der gleichen Nummer und unterschiedlichen Buchstaben (z. B. 3a., 3b., 3c.) gekennzeichnet.

Nachfolgend wird die erste Ergebnistabelle wiederum ausführlich beschrieben, um einmalig die vollständige Lesart der Tabelle zu erklären. Im Anschluss werden zur Reduktion der Darstellung jeweils nur die für die Beantwortung der Forschungsfrage wichtigsten Koeffizienten und Parameter der Tabellen erläutert.

### ***Schulische Selbstwirksamkeitserwartung***

In Tabelle 21 ist der modulspezifische Einfluss der wahrgenommenen Implementation auf die schulische Selbstwirksamkeitserwartung dargestellt. Zunächst fällt auf, dass die Intra-klassen-Korrelation Null beträgt (siehe Tabellenanmerkungen) und somit darauf hinweist, dass die Methode der Mehrebenenanalyse an dieser Stelle nicht unbedingt notwendig ist, sondern auch das Verfahren der Kovarianzanalyse zum Einsatz kommen könnte. Zur Vereinheitlichung des Vorgehens und aus Gründen der Vollständigkeit wird auch bei sehr kleinen oder gegen Null weisenden ICCs die Mehrebenenanalyse als Methode beibehalten. Hinsichtlich der Kriterien zur Beurteilung der Modellgüte zeigt sich im Vergleich zum Nullmodell (Devianz = 184.180; AIC = 190.180) beim autoregressiven Modell eine signifikant geringere Devianz (171.335) sowie ein kleinerer AIC (179.335). Das autoregressive Modell ist somit als das bessere Modell zu bewerten. Von den drei miteinander konkurrierenden Zielmodellen 3a (Devianz = 169.432; AIC = 179.432), 3b (Devianz = 171.163; AIC = 181.163) und 3c (Devianz = 161.234; AIC = 171.234), die zur Beantwortung der Fragestellung herangezogen werden, verfügt nur das Modell 3c über eine signifikant geringere Devianz und einen kleineren AIC als das autoregressive Modell. Dieses Implementationsmodell (T1, ImpS.llr), das den Einfluss der wahrgenommenen Umsetzung des Lern- und Leistungsraums auf die schulische Selbstwirksamkeitserwartung erfasst, stellt somit das beste Modell dar.

Das Nullmodell ohne jegliche Prädiktoren gibt noch keine Auskunft über den Einfluss der Implementation auf die Wirksamkeit und der Achsenabschnitt  $\gamma_{00}$  (2.911\*\*\*) stellt den Mittelwert aller Klassen zum zweiten Messzeitpunkt dar. Die Schülervarianzkomponente  $r_{ij}$  (0.252), die sich auf diesen Achsenabschnitt bezieht, soll in den nachfolgenden Modellen durch die Aufnahme der Implementationsprädiktoren erklärt werden. Wie bereits be-

schrieben, ist die Varianz zwischen den Klassen  $u_{0j}$  (0.000) für die Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler nicht von Interesse, da keine Prädiktoren auf der Klassenebene aufgenommen werden sollen. Durch die Wahl der Zentrierung (,Group-Mean-Zentrierung‘) wird diese aus dem Modell herausgerechnet und fällt entsprechend klein aus (Eid et al., 2013).

Im Anschluss an das Nullmodell werden im sogenannten autoregressiven Modell die Ausgangswerte in der schulischen Selbstwirksamkeitserwartung (hier gekennzeichnet mit ,T1‘) als Level-1-Prädiktor aufgenommen. Auch dieses Modell gibt noch keine Auskunft über den Einfluss der Implementation, sondern stellt einen Zwischenschritt vom Nullmodell zur Bildung der eigentlichen Zielmodelle dar. Da der Prädiktor ,T1‘, genau wie die in den späteren Modellen enthaltenen Implementationsprädiktoren (,ImpS.kk‘, ,ImpS.tp‘ und ,ImpS.llr‘), um die Klassenmittelwerte zentriert wird, beschreibt der Achsenabschnitt  $\gamma_{00}$  (2.911\*\*\*) den erwarteten Wert für Schüler aller Klassen zum zweiten Messzeitpunkt, deren Ausprägung in der Variablen ,T1‘ ihren jeweiligen Klassenmittelwerten entsprechen. Der Parameter  $\gamma_{10}$  (0.315\*\*\*) bildet den signifikanten Einfluss der Ausgangswerte auf die Werte zum zweiten Messzeitpunkt ab. Die Aufnahme der Ausgangswerte trägt vom Null- zum autoregressiven Modell zu einer Aufklärung der Schülervarianz von 9.9 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.099$ ;  $r_{ij} = 0.227$ ). Ein Prädiktor einer niedrigeren Ebene kann nicht nur zur Veränderung der Varianzkomponenten der eigenen Ebene, sondern auch zur Veränderung der Varianzanteile höherer Ebenen führen (Hox, 2010).<sup>54</sup> So erklärt sich der minimale Anstieg der Varianz zwischen Klassen  $u_{0j}$  (0.002) vom Null- zum Autoregressiven Modell. Diese Varianzkomponente wurde aus technischen Gründen modelliert, da in HLM mindestens ein Varianzparameter pro Ebene variieren muss, damit die Modelle iterieren und konvergieren können. Die Varianz  $u_{1j}$  wird, wie in der Modellbildung dargestellt (siehe 6.5.4), nicht modelliert.

Erst die Modelle 3a bis 3c zeigen den Einfluss der modulspezifischen wahrgenommenen Implementation seitens der Schüler auf die schulische Selbstwirksamkeitserwartung. Neben der Kovariaten ,T1‘ werden in diesen Modellen die Implementationsprädiktoren (,ImpS.kk‘, ,ImpS.tp‘ und ,ImpS.llr‘) aufgenommen. Der Parameter  $\gamma_{00}$  (3a-3c: 2.911\*\*\*) stellt wiederum den Achsenabschnitt dar und erfasst wegen der ,Group-Mean-Zentrierung‘

---

<sup>54</sup> Es sei daran erinnert, dass umgekehrt ein Prädiktor einer höheren Ebene nicht zu Veränderungen der Varianzanteile niedrigerer Ebenen beitragen kann (Hox, 2010). Deshalb bleiben die Schüler-Varianzkomponenten der linearen und diskontinuierlichen Modelle bis auf minimale, durch die Schätzverfahren verursachte Abweichungen unverändert.

der Prädiktoren den erwarteten Wert zum zweiten Messzeitpunkt für die Interventionsschüler, deren Ausprägungen in der Variablen ‚T1‘ und den Implementationsprädiktoren ihren jeweiligen Klassenmittelwerten entsprechen. Unter Kontrolle der Ausgangswerte ( $\gamma_{10} = 0.294^{***}/0.314^{***}/0.255^{***}$ ) beschreiben die Koeffizienten  $\gamma_{20(kk)}$  (0.120),  $\gamma_{20(tp)}$  (0.036) und  $\gamma_{20(llr)}$  (0.286\*\*) den Einfluss der wahrgenommenen modulspezifischen Implementation auf die schulische Selbstwirksamkeitserwartung zum zweiten Messzeitpunkt. Weder die wahrgenommene Umsetzung der Förderstrategien des Moduls Klassenklima noch die des Moduls Transparenz haben einen signifikanten Einfluss. Entsprechend tragen sie auch zu einer eher geringen Varianzaufklärung bei (Pseudo- $R^2$  von  $r_{ij} = 0.018/0.004$ ;  $r_{ij} = 0.223/0.226$ ). Die wahrgenommene Umsetzung der Modulinhalte zum Lern- und Leistungsraum hat jedoch einen signifikanten Einfluss. Unter Kontrolle der Ausgangswerte gilt: Je mehr Umsetzung der Modulinhalte zum Lern- und Leistungsraum von den Schülern wahrgenommen wird, desto höher ist ihre schulische Selbstwirksamkeitserwartung am Ende des ersten Projektjahres.<sup>55</sup> Der Implementationsprädiktor zum Lern- und Leistungsraum (‚ImpS.llr‘) trägt vom autoregressiven zum Implementationsmodell 3c zu einer Aufklärung der Varianz zwischen den Schülern von 8.4 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.084$ ;  $r_{ij} = 0.208$ ).

Zusammenfassend lässt sich festhalten: Unter Kontrolle der Ausgangswerte zeigt sich zum zweiten Messzeitpunkt kein bedeutsamer Einfluss der wahrgenommenen Implementation der Module Klassenklima und Transparenz auf die schulische Selbstwirksamkeitserwartung. Die von den Schülern wahrgenommene Umsetzung der Förderstrategien des Moduls Lern- und Leistungsraum hingegen hat unter Kontrolle der Ausgangswerte einen bedeutsamen Einfluss auf die schulische Selbstwirksamkeitserwartung am Ende des ersten Projektjahres. Entsprechend weist das Implementationsmodell 3c den besten Modellfit auf.

---

<sup>55</sup> Anders ausgedrückt bedeutet dies beispielsweise für zwei Schüler derselben Klasse mit gleichen Ausgangswerten, die sich jedoch hinsichtlich der wahrgenommenen Implementation bezogen auf das Modul zum Lern- und Leistungsraum um eine Skaleneinheit unterscheiden, dass der geschätzte Wert in der schulischen Selbstwirksamkeitserwartung für den Schüler mit der höheren Implementationswahrnehmung um 0.286 Skaleneinheiten über dem seines Klassenkameraden liegt.

**Tabelle 21:** Einfluss der wahrgenommenen Implementation auf die schulische Selbstwirksamkeitserwartung

n L1 = 126 n L2 = 10		1.	2.	3a.	3b.	3c.
Parameter		Nullmodell	Autoregressives Modell (T1)	Implementations- modell (T1, ImpS.kk)	Implementations- modell (T1, ImpS.tp)	Implementations- modell (T1, ImpS.llr)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	2.911*** (0.04)	2.911*** (0.04)	2.911*** (0.04)	2.911*** (0.04)	2.911*** (0.04)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.315*** (0.09)	0.294*** (0.09)	0.314*** (0.09)	0.255** (0.08)
Einfluss der wahrgenommenen Implementation – KK	$\gamma_{20(kk)}$	--	--	0.120 (0.09)	--	--
Einfluss der wahrgenommenen Implementation – TP	$\gamma_{20(tp)}$	--	--	--	0.036 (0.09)	--
Einfluss der wahrgenommenen Implementation – LLR	$\gamma_{20(llr)}$	--	--	--	--	0.286** (0.09)
<b>Zufallseffekte (Varianzkomponenten)</b>						
Level 1:	$r_{ij}$	0.252	0.227	0.223	0.226	0.208
Level 2:	$u_{0j}$	0.000	0.002	0.002	0.002	0.003
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>						
Level 1:	$r_{ij}$	--	0.099	0.018	0.004	0.084
<b>Modellfit</b>						
Devianz(q)		184.180(3)	171.335(4)	169.432(5)	171.163(5)	161.234(5)
$\chi^2$ -Differenz		--	12.845***	1.903	0.172	10.101**
AIC		190.180	179.335	179.432	181.163	171.234

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 2; 0.000; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum



### ***Prüfungsängstlichkeit***

In Tabelle 22 ist der Einfluss der von den Schülern wahrgenommenen modulspezifischen Implementation auf die Prüfungsängstlichkeit dargestellt. Die ICC (0.002) ist sehr klein. Die Kriterien zur Beurteilung der Modellgüte beschreiben das autoregressive Modell als das am besten geeignete Modell zur Vorhersage der Daten. Verglichen mit den Kennwerten des Nullmodells (Devianz = 256.406; AIC = 262.406) weist das autoregressive Modell eine geringere Devianz (243.514) und einen niedrigeren AIC (251.514) auf. Die Differenz (12.892\*\*\*) der beiden Devianzen wird im  $\chi^2$ -Test signifikant und spricht somit ebenfalls für die Güte des Modells. Wider Erwarten verfügt keines der Implementationsmodelle über einen besseren Modellfit. Zwar sind die Devianzen der Implementationsmodelle 3a-3c (242.952/243.377/242.567) minimal kleiner als die des autoregressiven Modells. Die Differenzen der Devianzen (0.562/0.137/0.947) werden jedoch nicht signifikant und der AIC aller Implementationsmodelle (252.952/253.377/252.567) fällt sogar größer aus als der des autoregressiven Modells.

Unter Kontrolle der Ausgangswerte zeigt sich kein bedeutsamer Einfluss der Implementationsprädiktoren ( $\gamma_{20(kk)} = -0.095$ ;  $\gamma_{20(tp)} = -0.048$ ;  $\gamma_{20(llr)} = 0.137$ ) auf die Prüfungsängstlichkeit. Zwar weisen die Koeffizienten bezogen auf die Implementationswahrnehmung der Module Klassenklima und Transparenz mit ihren negativen Vorzeichen in die erwartete Richtung: Je höher die wahrgenommene Implementation der Module, desto geringer fällt die Prüfungsängstlichkeit der Interventionsschüler unter Kontrolle der Ausgangswerte am Ende des ersten Projektjahres aus. Die Koeffizienten sind jedoch nicht signifikant. Auch der Parameter  $\gamma_{20(llr)}$ , der zum zweiten Messzeitpunkt unter Kontrolle der Ausgangswerte wider Erwarten einen tendenziellen Anstieg der Prüfungsängstlichkeit bei hoher wahrgenommener Implementation des Moduls Lern- und Leistungsraum beschreibt, ist nicht signifikant. Die Aufklärungsrate der Schüler-Varianz liegt verglichen mit dem autoregressiven Modell bei allen Implementationsmodellen unter 1 % (Pseudo- $R^2$  von  $r_{ij} = 0.006/0.002/0.008$ ).

Zusammenfassend lässt sich festhalten, dass unter Kontrolle der Ausgangswerte weder die von den Schülern wahrgenommene Umsetzung des Moduls Klassenklima noch die der Module Transparenz oder Lern- und Leistungsraum einen bedeutsamen Einfluss auf die Prüfungsängstlichkeit zum zweiten Messzeitpunkt haben. Die Kriterien zur Beurteilung der Modellgüte beschreiben das autoregressive Modell als das beste Modell.

**Tabelle 22:** Einfluss der wahrgenommenen Implementation auf die Prüfungsängstlichkeit

n L1 = 117 n L2 = 10	Parameter	1. Nullmodell	2. Autoregressives Modell (T1)	3a. Implementations- modell (T1, ImpS.kk)	3b. Implementations- modell (T1, ImpS.tp)	3c. Implementations- modell (T1, ImpS.llr)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	2.652***(0.07)	2.652***(0.07)	2.652***(0.07)	2.652***(0.07)	2.652***(0.07)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.351***(0.09)	0.356***(0.09)	0.354***(0.09)	0.349***(0.09)
Einfluss der wahrgenommenen Imple- mentation – KK	$\gamma_{20(kk)}$	--	--	-0.095(0.13)	--	--
Einfluss der wahrgenommenen Imple- mentation – TP	$\gamma_{20(tp)}$	--	--	--	-0.048(0.13)	--
Einfluss der wahrgenommenen Imple- mentation – LLR	$\gamma_{20(llr)}$	--	--	--	--	0.137(0.14)
<b>Zufallseffekte (Varianzkomponenten)</b>						
Level 1:	$r_{ij}$	0.532	0.474	0.471	0.473	0.470
Level 2:	$u_{0j}$	0.001	0.003	0.003	0.003	0.003
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>						
Level 1:	$r_{ij}$	--	0.109	0.006	0.002	0.008
<b>Modellfit</b>						
Devianz(q)		256.406(3)	243.514(4)	242.952(5)	243.377(5)	242.567(5)
$\chi^2$ -Differenz		--	12.892***	0.562	0.137	0.947
AIC		262.406	251.514	252.952	253.377	252.567

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 2; 0.002; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum

### ***Schulbezogenes Stresserleben – wahrgenommene Herausforderung***

Tabelle 23 gibt Auskunft über den Einfluss der von den Schülern wahrgenommenen modulspezifischen Implementation auf das schulbezogene Herausforderungserleben. Die ICC ist mit 0.205 nach Hox (2010) als sehr groß zu bewerten. Nach den Kriterien zur Beurteilung der Modellgüte weist das Implementationsmodell 4b, in dem die wahrgenommene Implementation der Förderstrategien des Moduls Klassenklima und des Moduls Lern- und Leistungsraum berücksichtigt wird, den besten Modellfit auf. Verglichen mit den Kennwerten des Null- (Devianz = 255.644; AIC = 261.644) und des autoregressiven Modells (Devianz = 234.612; AIC = 242.612) sowie der Implementationsmodelle 3 und 4a (Devianz = 227.696/225.361; AIC = 237.696/237.361) verfügt das Modell 4b über die kleinste Devianz (220.164) und den geringsten AIC (232.164). Auch der  $\chi^2$ -Test spricht für die Güte des Modells, denn die Differenz der Devianzen (7.532\*\*) der ineinander geschachtelten Modelle 3 und 4b wird signifikant. Im Gegensatz dazu wird die Differenz der Devianzen (2.335) der alternativ geschachtelten Modelle 3 und 4a nicht signifikant und die Devianz sowie der AIC der beiden Modelle unterscheiden sich kaum (s. o.). Das Implementationsmodell 4a, in das die Implementationsindikatoren Klassenklima und Transparenz aufgenommen werden, ist folglich weniger gut zur Vorhersage der Daten geeignet als die Implementationsmodelle 3 und 4b.

Entsprechend den Kriterien zur Beurteilung der Modellgüte, die das Implementationsmodell 3 als ein gutes Modell charakterisieren und das Modell 4b als das beste Modell auszeichnen, zeigt sich unter Kontrolle der Ausgangswerte in beiden Modellen ein signifikanter Einfluss der jeweils modellierten Implementationsindikatoren: Zunächst wird im Modell 3 der Koeffizient  $\gamma_{20(kk)}$  (0.275\*\*), der den Einfluss der wahrgenommenen Implementation der Förderstrategien des Moduls Klassenklima auf das Herausforderungserleben beschreibt, signifikant. Die Aufnahme des Prädiktors ‚ImpS.kk‘ trägt vom autoregressiven Modell 2 zum Implementationsmodell 3 zu einer Varianzaufklärung von 5.7 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.057$ ). Da der Prädiktor ‚ImpS.kk‘ signifikant ist, verbleibt er im Modell und zusätzlich wird der chronologisch nachfolgende Implementationsindikator ‚ImpS.tp‘ im Modell 4a modelliert. In diesem Modell ist der Koeffizient  $\gamma_{20(kk)}$  (0.239\*) weiterhin signifikant. Der Parameter  $\gamma_{30(tp)}$  (0.164), der den Einfluss der wahrgenommenen Implementation des Moduls Transparenz auf das Herausforderungserleben erfasst, weist zwar durch sein positives Vorzeichen in die erwartete Richtung (je mehr Transparenz wahrgenommen wird, desto höher ist das Herausforderungserleben in der Schule). Der Effekt ist

jedoch nicht signifikant und die zusätzliche Aufnahme des Prädiktors ‚ImpS.tp‘ trägt vom Implementationsmodell 3 zum Modell 4a zu einer Varianzaufklärung von 2.2 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.022$ ). Da der Prädiktor ‚ImpS.tp‘ keinen bedeutsamen Beitrag zur Verbesserung der Modellgüte leistet (s. o.), wird er in den folgenden Modellen nicht weiter berücksichtigt. In das nächste Modell 4b wird stattdessen neben den Ausgangswerten und dem Implementationsindikator Klassenklima der Prädiktor ‚ImpS.llr‘ neu aufgenommen. Der zugehörige Koeffizient  $\gamma_{20(\text{llr})}$  (0.307\*), der den Einfluss der wahrgenommenen Implementation der Förderstrategien des Moduls Lern- und Leistungsraum auf das Herausforderungserleben erfasst, wird ebenso wie der Koeffizient  $\gamma_{20(\text{kk})}$  (0.203\*) signifikant. Dies bedeutet unter Kontrolle der Ausgangswerte für die Interventionsschüler: Je höher die wahrgenommene Implementation der Module Klassenklima und Lern- und Leistungsraum, desto größer wird das wahrgenommene Herausforderungserleben. Die zusätzliche Aufnahme des Prädiktors ‚ImpS.llr‘ trägt vom Modell 3 zum Modell 4b zu einer Varianzaufklärung von 6.4 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.064$ ).

Zusammenfassend lässt sich festhalten: Unter Kontrolle der Ausgangswerte hat sowohl die von den Schülern wahrgenommene Umsetzung des Moduls Klassenklima als auch die des Moduls Lern- und Leistungsraum einen bedeutsamen positiven Einfluss auf das Herausforderungserleben zum zweiten Messzeitpunkt. Die wahrgenommene Implementation des Moduls Transparenz hingegen trägt nicht zu einer signifikanten Veränderung des Herausforderungserlebens bei. Nach den Kriterien zur Beurteilung der Modellgüte weist das Implementationsmodell 4b mit den Implementationsprädiktoren Klassenklima und Transparenz den besten Modellfit auf.

**Tabelle 23:** Einfluss der wahrgenommenen Implementation auf das schulbezogene Herausforderungserleben

n L1 = 127 n L2 = 10		1.	2.	3.	4a.	4b.
Parameter		Nullmodell	Autoregressives Modell (T1)	Implementations- modell (T1, ImpS.kk)	Implementations- modell (T1, ImpS.kk&tp)	Implementations- modell (T1, ImpS.kk&llr)
<i>Feste Effekte</i>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	2.376*** (0.12)	2.375*** (0.12)	2.375*** (0.12)	2.375*** (0.12)	2.375*** (0.12)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.384*** (0.08)	0.331*** (0.08)	0.316*** (0.08)	0.289*** (0.08)
Einfluss der wahrgenommenen Implementation – KK	$\gamma_{20(kk)}$	--	--	0.275** (0.10)	0.239* (0.10)	0.203* (0.10)
Einfluss der wahrgenommenen Implementation – TP	$\gamma_{30(tp)}$	--	--	--	0.164 (0.11)	--
Einfluss der wahrgenommenen Implementation – LLR	$\gamma_{30(llr)}$	--	--	--	--	0.307** (0.11)
<i>Zufallseffekte (Varianzkomponenten)</i>						
Level 1:	$r_{ij}$	0.397	0.332	0.313	0.306	0.293
Level 2:	$u_{0j}$	0.102***	0.108***	0.110***	0.111***	0.112***
<i>Varianzaufklärung (Pseudo-<math>R^2</math>)</i>						
Level 1:	$r_{ij}$	--	0.164	0.057	0.022	0.064
Modellfit						
Devianz(q)		255.644(3)	234.612(4)	227.696(5)	225.361(6)	220.164(6)
$\chi^2$ -Differenz		--	21.032***	6.916**	2.335	7.532**
AIC		261.644	242.612	237.696	237.361	232.164

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 2 = 0.205; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum

***Schulbezogenes Stresserleben – wahrgenommene Bedrohung***

Tabelle 24 beschreibt den Einfluss der von den Schülern wahrgenommenen modulspezifischen Implementation auf das schulbezogene Bedrohungserleben. Die ICC weist mit 0.0004 wiederum gegen Null und ist als sehr klein zu bewerten (Hox, 2010). Nach den Kriterien zur Beurteilung der Modellgüte ist das autoregressive Modell das am besten geeignete Modell zur Vorhersage der Daten. Im Vergleich mit den Kennwerten des Nullmodells (Devianz = 266.851; AIC = 272.851) weist das autoregressive Modell eine geringere Devianz (224.737) und einen niedrigeren AIC (232.737) auf. Auch die Differenz (42.114\*\*\*) der beiden Devianzen wird im  $\chi^2$ -Test signifikant und ist somit ein weiteres Indiz für die Güte des Modells. Wider Erwarten verfügt keines der Implementationsmodelle über einen besseren Modellfit. Die Devianzen (224.673/224.733) und die AICs (234.673/234.733) der Implementationsmodelle 3a und 3c unterscheiden sich in ihrer Höhe nur geringfügig von den Kennwerten des autoregressiven Modells und die zugehörigen Differenzen der Devianzen werden nicht signifikant (0.064/0.004). Die Kennwerte des Implementationsmodells 3b (Devianz = 231.158; AIC = 241.158) sind sogar größer als die des autoregressiven Modells und die Differenz der beiden Modelle (-6.421\*) wird negativ und signifikant. Dies spricht sogar für eine bedeutsam schlechtere Modellgüte des Implementationsmodells 3b.

Unter Kontrolle der Ausgangswerte zeigt sich kein bedeutsamer Einfluss der Implementationsprädiktoren ( $\gamma_{20(kk)} = 0.027$ ;  $\gamma_{20(tp)} = 0.111$ ;  $\gamma_{20(lir)} = -0.007$ ) auf das schulbezogene Bedrohungserleben. Die Koeffizienten bezogen auf die Implementationswahrnehmung der Module Klassenklima und Transparenz weisen wider Erwarten sogar positive Vorzeichen auf, sie sind jedoch nicht signifikant. Auch der Parameter  $\gamma_{20(lir)}$ , der zum zweiten Messzeitpunkt unter Kontrolle der Ausgangswerte entsprechend der Erwartung eine minimale, gegen Null weisende Reduktion des Bedrohungserlebens bei hoher wahrgenommener Implementation des Moduls Lern- und Leistungsraum beschreibt, ist nicht signifikant. Das Pseudo- $R^2$  der Schüler-Varianzkomponente der Modelle 3a und 3c beträgt 0 %. Die Aufnahme der Implementationsprädiktoren Klassenklima und Lern- und Leistungsraum trägt folglich zu keiner Aufklärung der Schüler-Varianz bei. Beim Implementationsmodell 3b wird das Pseudo- $R^2$  der Schüler-Varianzkomponenten negativ und kann nicht interpretiert werden.

Zusammenfassend lässt sich festhalten, dass unter Kontrolle der Ausgangswerte weder die von den Schülern wahrgenommene Umsetzung des Moduls Klassenklima noch die der

Module Transparenz oder Lern- und Leistungsraum einen bedeutsamen Einfluss auf das schulbezogene Bedrohungserleben zum zweiten Messzeitpunkt haben. Die Kriterien zur Beurteilung der Modellgüte beschreiben das autoregressive Modell als das beste Modell.

**Tabelle 24:** Einfluss der wahrgenommenen Implementation auf das schulbezogene Bedrohungserleben

n L1 = 117 n L2 = 10		1.	2.	3a.	3b.	3c.
Parameter		Nullmodell	<b>Autoregressives Modell (T1)</b>	Implementationsmodell (T1, ImpS.kk)	Implementationsmodell (T1, ImpS.tp)	Implementationsmodell (T1, ImpS.llr)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	2.042*** (0.06)	2.036*** (0.06)	2.036*** (0.06)	2.035*** (0.07)	2.036*** (0.06)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.611*** (0.09)	0.610*** (0.09)	0.615*** (0.09)	0.610*** (0.09)
Einfluss der wahrgenommenen Implementation – KK	$\gamma_{20(kk)}$	--	--	0.027(0.10)	--	--
Einfluss der wahrgenommenen Implementation – TP	$\gamma_{20(tp)}$	--	--	--	0.111(0.11)	--
Einfluss der wahrgenommenen Implementation – LLR	$\gamma_{20(llr)}$	--	--	--	--	-0.007(0.11)
<b>Zufallseffekte (Varianzkomponenten)</b>						
Level 1:	$r_{ij}$	0.487	0.338	0.338	0.340	0.338
Level 2:	$u_{0j}$	0.0002	0.013	0.013	0.018	0.013
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>						
Level 1:	$r_{ij}$	--	0.306	0.000	-0.006	0.000
<b>Modellfit</b>						
Devianz(q)		266.851(3)	224.737(4)	224.673(5)	231.158(5)	224.733(5)
$\chi^2$ -Differenz		--	42.114***	0.064	-6.421*	0.004
AIC		272.851	232.737	234.673	241.158	234.733

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 2 = 0.0004; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum



### ***Schulbezogenes Stresserleben – wahrgenommener Verlust***

In Tabelle 25 ist der Einfluss der von den Schülern wahrgenommenen modulspezifischen Implementation auf das schulbezogene Verlusterleben dargestellt. Die ICC ist mit 0.022 nach Hox (2010) als klein zu bewerten. Bei Betrachtung der Kriterien zur Beurteilung der Modellgüte erweist sich erneut das autoregressive Modell als das am besten geeignete Modell zur Vorhersage der Daten. Verglichen mit den Kennwerten des Nullmodells (Devianz = 256.570; AIC = 262.570) weist das autoregressive Modell eine geringere Devianz (249.040) und einen niedrigeren AIC (257.040) auf. Auch die signifikante Differenz (7.530\*\*) der beiden Devianzen spricht für die Güte des Modells. Keines der Implementationsmodelle verfügt über einen besseren Modellfit. Die Devianzen der Implementationsmodelle 3a-3c (248.537/247.776/247.704) sind nur minimal kleiner als die des autoregressiven Modells und die Differenzen zwischen der Devianz des jeweiligen Modells 3a-c und der Devianz des autoregressiven Modells (0.503/1.264/1.336) werden nicht signifikant. Der AIC aller Implementationsmodelle (258.537/257.776/257.704) fällt verglichen mit dem AIC des autoregressiven Modells (s. o.) sogar geringfügig größer bis identisch aus.

Unter Kontrolle der Ausgangswerte findet sich kein bedeutsamer Einfluss der Implementationsprädiktoren ( $\gamma_{20(kk)} = 0.082$ ;  $\gamma_{20(tp)} = 0.133$ ;  $\gamma_{20(llr)} = -0.146$ ) auf das schulbezogene Verlusterleben. Die Koeffizienten bezogen auf die Implementationswahrnehmung der Module Klassenklima und Transparenz verweisen in der Tendenz mit ihren positiven Vorzeichen unter Kontrolle der Ausgangswerte auf eine erwartungswidrige Zunahme des Verlusterlebens bei hoher wahrgenommener Implementation der Module. Jedoch sind beide Parameter nicht signifikant und die Aufklärungsrate der Schüler-Varianz liegt verglichen mit dem autoregressiven Modell bei beiden Implementationsmodellen 3a und 3b bei unter beziehungsweise gleich 1 % (Pseudo- $R^2$  von  $r_{ij} = 0.002/0.010$ ). Auch der Parameter  $\gamma_{20(llr)}$ , der durch sein negatives Vorzeichen zum zweiten Messzeitpunkt unter Kontrolle der Ausgangswerte erwartungskonform eine tendenzielle Reduktion des Verlusterlebens bei hoher wahrgenommener Implementation des Moduls Lern- und Leistungsraum beschreibt, ist nicht signifikant. Die Aufnahme des Prädiktors ‚ImpS.llr‘ trägt verglichen mit dem autoregressiven Modell ebenfalls zu einer Varianzaufklärung von 1 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.010$ ).

Zusammenfassend lässt sich festhalten, dass unter Kontrolle der Ausgangswerte die von den Schülern wahrgenommene Implementation der Modulinhalte des ersten Projektjahres

## Ergebnisse

keinen bedeutsamen Einfluss auf das Verlusterleben zum zweiten Messzeitpunkt hat. Nach den Kriterien zur Beurteilung der Modellgüte ist das autoregressive Modell das am besten geeignete Modell zur Vorhersage der Daten.

**Tabelle 25:** Einfluss der wahrgenommenen Implementation auf das schulbezogene Verlusterleben

n L1 = 127 n L2 = 10		1.	2.	3a.	3b.	3c.
Parameter		Nullmodell	Autoregressives Modell (T1)	Implementations- modell (T1, ImpS.kk)	Implementations- modell (T1, ImpS.tp)	Implementations- modell (T1, ImpS.llr)
<b>Feste Effekte</b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	1.753*** (0.07)	1.753*** (0.07)	1.753*** (0.07)	1.753*** (0.07)	1.753*** (0.07)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.242** (0.09)	0.236** (0.09)	0.252** (0.09)	0.223* (0.09)
Einfluss der wahrgenommenen Implementation – KK	$\gamma_{20(kk)}$	--	--	0.082 (0.11)	--	--
Einfluss der wahrgenommenen Implementation – TP	$\gamma_{20(tp)}$	--	--	--	0.133 (0.12)	--
Einfluss der wahrgenommenen Implementation – LLR	$\gamma_{20(llr)}$	--	--	--	--	-0.146 (0.13)
<b>Zufallseffekte (Varianzkomponenten)</b>						
Level 1:	$r_{ij}$	0.439	0.411	0.410	0.407	0.407
Level 2:	$u_{0j}$	0.010	0.013	0.013	0.013	0.013
<b>Varianzaufklärung (Pseudo-<math>R^2</math>)</b>						
Level 1:	$r_{ij}$	--	0.064	0.002	0.010	0.010
<b>Modellfit</b>						
Devianz(q)		256.570(3)	249.040(4)	248.537(5)	247.776(5)	247.704(5)
$\chi^2$ -Differenz		--	7.530**	0.503	1.264	1.336
AIC		262.570	257.040	258.537	257.776	257.704

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; ICC für Level 2 = 0.022; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum

***Binnendifferenzierung***

Tabelle 26 gibt Auskunft über den Einfluss der von den Schülern wahrgenommenen modulspezifischen Implementation auf die erlebte Binnendifferenzierung im Unterricht. Nach den Kriterien zur Beurteilung der Modellgüte weist das Implementationsmodell 4, in dem die wahrgenommene Implementation der Förderstrategien der Module Transparenz und Lern- und Leistungsraum berücksichtigt wird, den besten Modellfit auf. Im Vergleich zu den Kennwerten des Null- (Devianz = 231.942; AIC = 237.942) und des autoregressiven Modells (Devianz = 230.810; AIC = 238.810) sowie der Implementationsmodelle 3a und 3b (Devianz = 230.635/222.276; AIC = 240.635/232.276) verfügt das Modell 4 über die kleinste Devianz (217.456) und den geringsten AIC (229.456). Der  $\chi^2$ -Test spricht ebenfalls für die Güte des Modells, denn die Differenz der Devianzen (4.82\*) der ineinander geschachtelten Modelle 3b und 4 wird signifikant. Für diesen Differenzvergleich wird das Implementationsmodell 3b mit dem Prädiktor ‚ImpS.tp‘ herangezogen, weil das Modell 3b eine bessere Vorhersagefähigkeit aufweist als das alternative Vergleichsmodell 3a mit dem Prädiktor ‚ImpS.kk‘. Denn sowohl die Devianz als auch der AIC des Modells 3b sind geringer als die entsprechenden Kennwerte des Modells 3a (s.o.). Zudem wird die Differenz zwischen dem autoregressiven Modell und dem Modell 3b (8.534\*\*) im Gegensatz zur Differenz zwischen dem autoregressiven Modell und dem Modell 3a (0.175) signifikant. Das Modell 3a ist folglich schlechter zur Vorhersage der Daten geeignet als das Modell 3b, das von allen Modellen den höchsten AIC aufweist (s.o.).<sup>56</sup>

Entsprechend den Kriterien zur Beurteilung der Modellgüte, die das Implementationsmodell 3b als ein gutes Modell charakterisieren und das Modell 4 als das beste Modell hervorheben, zeigt sich unter Kontrolle der Ausgangswerte in beiden Modellen ein signifikanter Einfluss der jeweils modellierten Implementationsindikatoren. Nachdem sich der Prädiktor ‚ImpS.kk‘ mit  $\gamma_{20(kk)}$  (0.051) im Modell 3a als nicht signifikant erwiesen hat und das Modell von allen Modellen die schlechteste Modellgüte aufweist (s. o.), wird der Implementationsindikator Klassenklima verworfen und im Modell 3b zunächst der Koeffizient  $\gamma_{20(tp)}$  (0.342\*\*) modelliert. Dieser besagt, dass die wahrgenommene Implementation der Förderstrategien des Moduls Transparenz unter Kontrolle der Ausgangswerte einen

---

<sup>56</sup> Obwohl das autoregressive Modell keine bessere Modellgüte als das Nullmodell aufweist (der AIC des autoregressiven Modells ist mit 238.810 sogar höher als der des Nullmodells mit 237.942 und die Differenz der Devianzen beider Modelle mit 1.132 wird nicht signifikant), wird das autoregressive Modell als Vergleichsmodell für die Implementationsmodelle verwendet, um so die kovarianzanalytische Logik aufrechtzuerhalten. Die schlechtere Modellgüte des autoregressiven Modells ist vermutlich auf den nicht signifikanten Zusammenhang zwischen den Ausgangswerten und den Werten zum zweiten Messzeitpunkt zurückführbar. Möglicherweise hängt diese auch mit der geringen Reliabilität der Skala (siehe Anhang B Skalenhandbuch) zusammen. Die Ergebnisse müssen entsprechend mit Vorsicht interpretiert werden.

signifikanten Einfluss auf die erlebte Binnendifferenzierung zum zweiten Messzeitpunkt hat: Je mehr Transparenz wahrgenommen wird, desto höher ist auch die erlebte Binnendifferenzierung. Die Aufnahme des Prädiktors ‚ImpS.tp‘ trägt vom autoregressiven Modell 2 zum Implementationsmodell 3b zu einer Varianzaufklärung von 7.5 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.075$ ). Da der Prädiktor ‚ImpS.tp‘ signifikant ist, verbleibt er im Modell und zusätzlich wird der chronologisch nachfolgende Implementationsindikator ‚ImpS.llr‘ im Modell 4 modelliert (für eine ausführliche Erläuterung der Modellbildung siehe auch 6.5.4). In diesem Modell ist der Koeffizient  $\gamma_{20(tp)}$  (0.482\*\*\*) weiterhin signifikant. Der ebenfalls signifikante Parameter  $\gamma_{30(llr)}$  (-0.302\*), verweist durch sein negatives Vorzeichen unter Kontrolle der Ausgangswerte wider Erwarten auf einen negativen Zusammenhang zwischen der erlebten Binnendifferenzierung und der wahrgenommenen Implementation der Modulinhalt des Lern- und Leistungsraumes. Die zusätzliche Aufnahme des Prädiktors ‚ImpS.llr‘ trägt vom Implementationsmodell 3b zum Modell 4 zu einer Varianzaufklärung von 4.2 % bei (Pseudo- $R^2$  von  $r_{ij} = 0.042$ ).

Zusammenfassend lässt sich festhalten: Unter Kontrolle der Ausgangswerte hat die von den Schülern wahrgenommene Umsetzung des Moduls Transparenz einen bedeutsamen positiven Einfluss auf die erlebte Binnendifferenzierung zum zweiten Messzeitpunkt. Im Gegensatz dazu zeigt sich unter Kontrolle der Ausgangswerte und der Umsetzung des Moduls Transparenz ein negativer Einfluss der wahrgenommenen Umsetzung des Moduls Lern- und Leistungsraum. Nach den Kriterien zur Beurteilung der Modellgüte weist das Implementationsmodell 4 mit den Implementationsprädiktoren Transparenz und Lern- und Leistungsraum den besten Modellfit auf.

**Tabelle 26:** Einfluss der wahrgenommenen Implementation auf die wahrgenommene Binnendifferenzierung

n L1 = 119 n L2 = 10		1.	2.	3a.	3b.	4.
Parameter		Nullmodell	Autoregressives Modell (T1)	Implementations- modell (T1, ImpS.kk)	Implementations- modell (T1, ImpS.tp)	Implementations- modell (T1, ImpS.tp&llr)
<b><i>Feste Effekte</i></b>		Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)	Koeffizient (SE)
Achsenabschnitt	$\gamma_{00}$	2.274*** (0.06)	2.274*** (0.06)	2.274*** (0.06)	2.274*** (0.06)	2.274*** (0.06)
Kontrolle der Ausgangswerte	$\gamma_{10}$	--	0.091 (0.09)	0.082 (0.09)	0.097 (0.08)	0.090 (0.08)
Einfluss der wahrgenommenen Implementation – KK	$\gamma_{20(kk)}$	--	--	0.051 (0.12)	--	--
Einfluss der wahrgenommenen Implementation – TP	$\gamma_{20(tp)}$	--	--	--	0.342** (0.11)	0.482*** (0.13)
Einfluss der wahrgenommenen Implementation – LLR	$\gamma_{30(llr)}$	--	--	--	--	-0.302* (0.14)
<b><i>Zufallseffekte (Varianzkomponenten)</i></b>						
Level 1:	$r_{ij}$	0.415	0.411	0.410	0.380	0.364
Level 2:	$u_{0j}$	0.002	0.003	0.003	0.005	0.006
<b><i>Varianzaufklärung (Pseudo-<math>R^2</math>)</i></b>						
Level 1:	$r_{ij}$	--	0.010	0.002	0.075	0.042
Modellfit						
Devianz(q)		231.942(3)	230.810(4)	230.635(5)	222.276(5)	217.456(6)
$\chi^2$ -Differenz			1.132	0.175	8.534**	4.82*
AIC		237.942	238.810	240.635	232.276	229.456

Anmerkungen: n = Stichprobengröße; L1-2 = Level 1-2; SE = Standardschätzfehler; q = Anzahl der Parameter; \* p<.05; \*\* p<.01; \*\*\* p<.001; ICC für Level 2 = 0.005; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum

### **7.2.3 Zusammenfassung der Ergebnisse der Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation**

Hinsichtlich des Einflusses der wahrgenommenen modulspezifischen Implementation seitens der Schüler zeigen sich für das erste Projektjahr folgende signifikante Effekte: Je höher die wahrgenommene Umsetzung des Moduls Klassenklima, desto größer ist das Herausforderungserleben. Die Wahrnehmung der Umsetzung des zweiten Moduls hat einen positiven Einfluss auf die erlebte Binnendifferenzierung im Unterricht. Das heißt, je höher das Transparenzerleben der Schüler, desto größer ist auch die wahrgenommene Binnendifferenzierung. Die meisten Zusammenhänge mit den Zielkriterien weist die wahrgenommene Implementation des Moduls Lern- und Leistungsraum auf: positive Zusammenhänge mit der schulischen Selbstwirksamkeitserwartung und dem Herausforderungserleben sowie einen negativen Zusammenhang mit der erlebten Binnendifferenzierung. Eine ausführliche Interpretation dieser Ergebnisse erfolgt in der Diskussion (Kapitel 8).

## 8 Diskussion

Ziel der vorliegenden Arbeit war die Evaluation der Wirksamkeit des FOSIS2-Programms im Hinblick auf motivational relevante Ressourcen (schulische Selbstwirksamkeitserwartung, wahrgenommene Herausforderung und Binnendifferenzierung) und Risikofaktoren (Prüfungsängstlichkeit, wahrgenommene Bedrohung und wahrgenommener Verlust). Erwartet wurden bei den Interventionsschülern im Vergleich zu den Kontrollschülern über die beiden Projektjahre hinweg günstigere Entwicklungen in den benannten Variablen. Explorativ wurde außerdem geprüft, ob sich entgegen dieser linearen Verlaufsannahmen jahresspezifische, diskontinuierliche Effekte des FOSIS2-Programms ermitteln lassen. Darüber hinaus wurde beispielhaft für das erste Projektjahr der Zusammenhang zwischen der von den Schülern wahrgenommenen Implementation und der Programmwirksamkeit untersucht. Nachfolgend werden die Ergebnisse in Bezug zu den Fragestellungen und Hypothesen der Arbeit (siehe auch Kapitel 5) sowie die methodischen Einschränkungen der Studie diskutiert. Das Kapitel schließt mit einer Zusammenfassung und einem Ausblick.

### 8.1 Diskussion der Ergebnisse zur Wirksamkeitsevaluation

Im Hinblick auf die Wirksamkeit des FOSIS2-Programms wurde zunächst die kontinuierliche Entwicklung der Zielkriterien über beide Projektjahre hinweg mittels linearer Wachstumskurvenmodelle untersucht. Darüber hinaus gibt die Analyse der Veränderung mithilfe von diskontinuierlichen Mehrebenenmodellen Auskunft über mögliche jahresspezifische Effekte. Die Ergebnisse werden nachfolgend getrennt für die einzelnen Konstrukte diskutiert.

#### *Schulische Selbstwirksamkeitserwartung*

Bereits die grafischen Darstellungen der linearen und jahresspezifischen Verläufe (siehe Abbildung 11 und Abbildung 13) veranschaulichen, dass sich Interventions- und Kontrollgruppe in der Veränderung ihrer schulischen Selbstwirksamkeitserwartung nicht unterscheiden. In beiden Gruppen bleibt die Selbstwirksamkeitserwartung über den gesamten Projektzeitraum relativ stabil auf einem ähnlich hohen Niveau. Dieser bildliche Eindruck wird durch die Ergebnisse der Mehrebenenanalysen bestätigt: Die Veränderungsparameter der linearen und diskontinuierlichen Evaluationsmodelle werden nicht signifikant, das heißt, Interventions- und Kontrollgruppe unterscheiden sich entgegen der Hypothese nicht



bedeutsam in ihrer Selbstwirksamkeitserwartungsentwicklung.<sup>57</sup> Das Ergebnis entspricht entwicklungspsychologischen Erkenntnissen zur Selbstwirksamkeitserwartung, wonach diese als eine personale Ressource insbesondere ab dem Jugendalter zunehmend stabil bleibt und nur schwer zu beeinflussen ist (Wigfield et al., 2011). Es lässt sich vermuten, dass diese eher stabile, personale Ressource erst durch eine fortlaufende Implementation der FOSIS2-Maßnahmen hätte beeinflusst werden können. Die Analysen zum Implementationseinfluss zeigen zumindest für die Maßnahme Trennung von Lern- und Leistungsräumen einen Zusammenhang zwischen der von den Schülern wahrgenommenen Implementation und der Selbstwirksamkeitserwartung (siehe auch Abschnitt 8.2.2). Nach Patry und Hager (2000) ist eine wesentliche Schwierigkeit von Interventionsstudien, die auf die Verbesserung des Unterrichts abzielen, dass Veränderungen oft erst nach mehreren Jahren greifen und erst dann ihre Wirkungen entfalten. Wird im Rahmen einer klassischen Wirksamkeitsüberprüfung mittels Prä- und Post-Testvergleichen zunächst kein Effekt der Intervention, jedoch nach einiger Zeit bei Follow-Up-Testungen ein Effekt gefunden, wird dies auch als *„sleeper effect“* bezeichnet (Köller, 2009). Zu kurzfristige Evaluationsabstände können folglich dazu führen, dass die Effekte von Fortbildungsprogrammen zumindest teilweise nicht erfasst werden. Im FOSIS2-Projekt wurde versucht, diesem Problem zu begegnen, indem ein vergleichsweise langer Evaluationszeitraum von zwei Schuljahren gewählt wurde. Follow-Up-Untersuchungen wären wünschenswert gewesen, um die Nachhaltigkeit von bereits erfassten Effekten zu prüfen und zu untersuchen, ob sich bei stabilen Konstrukten wie der Selbstwirksamkeitserwartung in Folge einer lang anhaltenden, kontinuierlichen Implementation der FOSIS2-Maßnahmen ein möglicher *„sleeper effect“* nachweisen lässt.

### ***Prüfungsängstlichkeit***

Die Parameter der linearen Analysen veranschaulichen, dass die Prüfungsängstlichkeit in der Interventionsgruppe im Verlauf der beiden Projektjahre hypothesenkonform stärker abnimmt als in der Kontrollgruppe; der Gruppenunterschied in der Veränderung ist jedoch nicht signifikant. Da durch die Aufnahme des Prädiktors *„Gruppe“* aber 40 % der Varianz

---

<sup>57</sup> Angesichts der sehr kleinen ICC von 0.025 war dieser nicht vorhandene Effekt bereits nach der Modellierung des Nullmodells zu erwarten. Denn die ICC beschreibt den Anteil der Varianz zwischen den Klassen an der Gesamtvarianz (Hox, 2010). Wenn die Varianz zwischen den Klassen in dieser Studie sehr klein ausfällt, lässt sich dies im Hinblick auf die Evaluationsfragestellung bereits als ein Zeichen für einen nicht vorhandenen Effekt der Intervention werten. Denn entsprechend den Hypothesen der vorliegenden Arbeit werden Unterschiede zwischen den Interventions- und Kontrollklassen erwartet. Die Modellierung der Daten hätte folglich bereits nach der Bildung des Nullmodells beendet werden können. Aus Gründen der Vollständigkeit wurden jedoch auch für Nullmodelle mit einer sehr kleinen ICC die nachfolgenden Modelle mit Prädiktoren berechnet (siehe auch Anhang C1).

bezogen auf die Veränderung aufgeklärt werden kann, die zugehörige Varianzkomponente zwischen den Klassen im Evaluationsmodell nicht mehr signifikant wird und die Kriterien zur Beurteilung der Modellgüte das Evaluationsmodell als das beste Modell beschreiben (siehe Tabelle 7), ist die hypothesenkonforme Reduktion der Prüfungsangst in der Interventionsgruppe trotz nicht vorhandener Signifikanz als relevant zu bewerten. Die diskontinuierlichen Modelle weisen darauf hin, dass der Effekt vor allem auf das erste Projektjahr zurückzuführen ist, wo sich ein signifikanter Gruppenunterschied in der Veränderung nachweisen lässt: Die Prüfungsängstlichkeit nimmt in den Interventionsklassen im ersten Projektjahr signifikant stärker ab als in den Kontrollklassen. Im zweiten Projektjahr steigt die Prüfungsängstlichkeit in der Interventionsgruppe wieder minimal an (siehe Abbildung 14). Der Gruppenunterschied in der Veränderung wird im zweiten Projektjahr nicht signifikant, die Aufnahme des Prädiktors ‚Gruppe‘ führt hier aber immerhin zu einer Varianzaufklärung von 33 %. Die bedeutsame Reduktion der Prüfungsängstlichkeit im ersten Projektjahr lässt sich möglicherweise dadurch begründen, dass in diesem Zeitraum die eher angst- und belastungsreduzierenden FOSIS2-Module (wie die Maßnahmen zur Förderung des Klassenklimas, der Transparenz und des Lernraums) umgesetzt wurden. Ein positives Klassenklima sollte die Prüfungsängstlichkeit verringern (Eder, 2010) und das Transparenzpapier wurde explizit als Maßnahme zur Reduktion der Prüfungsängstlichkeit entwickelt (siehe z. B. Strittmatter, 1997). Auch der Lernraum soll Angst- und Belastungserleben reduzieren. Denn Lernräume sind Zeiten im Unterricht, in denen eine tolerante Fehlerkultur vorherrscht und die Schüler lernen und üben können, ohne angesichts möglicher Fehler eine schlechte Benotung befürchten zu müssen (Chott, 1999; 2004). Der Anstieg in der Prüfungsängstlichkeit der Interventionsgruppe im zweiten Projektjahr könnte damit zusammenhängen, dass die Umsetzung dieser entlastenden Module aufgrund der parallelen Einführung weiterer FOSIS2-Maßnahmen im zweiten Jahr wieder leicht zurückgegangen sein könnte. Nachdem die Interventionsschüler im ersten Jahr die positiven Auswirkungen einer angstfreien Lernkultur erlebt haben, lässt sich vermuten, dass sie im Kontrast hierzu das Ausbleiben oder eine Verringerung des Einsatzes der bereits eingeführten angstreduzierenden Maßnahmen umso deutlicher wahrnehmen, wodurch sich das Angsterleben im zweiten Jahr wieder verstärkt haben könnte. Um dies zu prüfen, hätte bei den modulspezifischen Rückmeldungen nicht nur die Umsetzung der Maßnahmen des aktuellen Moduls, sondern auch die Umsetzung der zuvor eingeführten Module miterhoben werden müssen. Dadurch wäre der Fragebogen der modulspezifischen Rückmeldung allerdings mit jedem neu eingeführten Modul länger ausgefallen und hätte aufgrund des Umfangs möglicherweise zunehmend Reaktanz

erzeugt. Deshalb wurde in der vorliegenden Studie darauf verzichtet, nach jedem Modul die Implementation aller bereits eingeführten FOSIS2-Maßnahmen zu erfassen.

### ***Schulbezogenes Stresserleben – wahrgenommene Herausforderung***

Bei den Ergebnissen zum Herausforderungserleben sind zunächst die signifikant unterschiedlichen Ausgangswerte zu beachten. Die wahrgenommene, schulbezogene Herausforderung wird von den Interventionsschülern zu Beginn des Projekts signifikant höher eingestuft als von den Kontrollschülern. Wo dieser Unterschied herrührt, lässt sich nur vermuten. Anders als die Kontrollschüler, die bei der ersten Befragung die Information erhielten, im Laufe der kommenden zwei Schuljahre noch weitere zwei Male an einer Befragung teilnehmen zu dürfen, in denen ihr schulisches Erleben und Aspekte des Unterrichtsgeschehens erfragt werden, war den Interventionsschülern darüber hinaus bekannt, dass das FOSIS2-Projekt nicht nur Befragungen, sondern auch Fortbildungen für ihre Lehrkräfte beinhaltet. Dieses Wissen um die FOSIS2-Maßnahmen und auch eventuelle Berichte ihrer Lehrer über das Programm könnten dazu geführt haben, dass die Interventionsschüler positive Erwartungen an die FOSIS2-Intervention entwickelt und bereits die Chance zur Teilnahme ihrer Lehrer am FOSIS2-Programm als positive Herausforderung bewertet haben. Die Ergebnisse zu den Gruppenunterschieden in der Entwicklung der wahrgenommenen Herausforderung können vor dem Hintergrund der unterschiedlichen Ausgangswerte nachfolgend nur mit Vorsicht interpretiert werden.

Im Verlauf der beiden Projektjahre verzeichnet die lineare Modellierung sowohl in der Interventions- als auch in der Kontrollgruppe einen Abfall im schulbezogenen Herausforderungserleben. Ohne den Einsatz der FOSIS2-Maßnahmen ist eine solche Reduktion der wahrgenommenen schulbezogenen Herausforderung insofern zu erwarten, als der Wechsel von der Grund- zur weiterführenden Schule mit bestimmten organisationalen und strukturellen Veränderungen einhergeht, die motivationale Ressourcen, wie das Herausforderungserleben, verringern können: Die meisten weiterführenden Schulen zeichnen sich beispielsweise durch größere Klassen aus als die Grundschulen, sodass eine Lehrperson gleichzeitig sehr viele Schüler beaufsichtigen muss, wodurch die Unterrichtsführung oft strenger wird. Zudem werden die Schüler auf den weiterführenden Schulen in der Regel von vielen verschiedenen Lehrkräften unterrichtet, wodurch die Entwicklung einer engen Schüler-Lehrer-Beziehung erschwert ist und sich die Schüler weniger von den Lehrkräften unterstützt fühlen. Neben diesen sozialen Veränderungen müssen die Schüler nach dem Wechsel auf die weiterführende Schule auch eine wachsende Anzahl an Aufgaben und

Anforderungen bewältigen (Wigfield et al., 2011) und kooperative Lernformen werden zunehmend von stärker kompetitiv ausgerichteten Lernsituationen abgelöst (Schiefele, 2009). Diese strukturellen Veränderungen verringern das Gefühl von Zugehörigkeit und Eingebundenheit der Schüler, wodurch motivationale Ressourcen abnehmen (Wigfield et al., 2011; Wigfield et al., 2012). Die FOSIS2-Maßnahmen sollten dem entgegenwirken und unter anderem dazu beitragen, dass sich die wahrgenommene Herausforderung bei den Interventionsschülern im Vergleich zu den Kontrollschülern günstiger entwickelt. Entgegen dieser Hypothese zeigt sich im Rahmen der linearen Modellierung in der Interventionsgruppe im Verlauf der beiden Projektjahre jedoch eine signifikant größere Reduktion des schulbezogenen Herausforderungserlebens als in der Kontrollgruppe. Zwar verliert sich dieser Effekt im Rahmen der diskontinuierlichen Analysen, wo sich in beiden Projektjahren keine signifikanten Gruppenunterschiede in der Veränderung zeigen. Aber auch hier weisen die Parameter insbesondere im ersten Projektjahr bei den Interventionsschülern auf eine stärkere Abnahme in der wahrgenommenen Herausforderung hin als bei den Kontrollschülern und die Aufnahme des Prädiktors ‚Gruppe‘ trägt immerhin zu einer Varianzaufklärung von 33 % bei. Diese stärkere Verringerung des Herausforderungserlebens bei den Interventionsschülern im ersten Jahr lässt sich eventuell ebenfalls dadurch erklären, dass in diesem Zeitraum vor allem die angst- und belastungsreduzierenden FOSIS2-Module eingeführt wurden. Diese sollen Ängstlichkeit und Bedrohungserleben reduzieren und könnten zu der Haltung verleiten, „die Schule etwas leichter nehmen zu können“, wodurch gleichzeitig möglicherweise zunächst der Eindruck entsteht, weniger herausgefordert zu werden. Vorzustellen ist, dass sich die wahrgenommene Herausforderung erst nach einer langfristigen stetigen Implementation einer angst- und bedrohungsfreien Fehlerkultur positiv verändert und die Schüler erst zeitverzögert mutiger werden, Herausforderungen zu suchen und schulbezogene Anforderungen günstig zu bewerten. Die Analysen zum Implementations-einfluss weisen bei den Interventionsschülern zumindest für die Umsetzung der Maßnahmen zum Klassenklima und dem Lern- und Leistungsraum auf einen Zusammenhang zwischen der wahrgenommenen Implementation und dem Herausforderungserleben hin (siehe auch Abschnitt 8.2.2). Eine weitere mögliche Begründung für die stärkere Reduktion des Herausforderungserlebens in der Interventionsgruppe wäre, dass diese auf zu hohe Erwartungen an die FOSIS2-Maßnahmen zurückzuführen ist. So könnte es sein, dass die Schüler sich von den Fortbildungen für ihre Lehrkräfte (noch) andere oder intensivere Veränderungen des Unterrichtsverhaltens versprochen als die tatsächlich erzeugten. Die Enttäuschung

über die Nicht-Erfüllung der Erwartungen könnte das Gefühl verstärkt haben, in der Schule nicht genügend positive Herausforderungen zu erleben.

### ***Schulbezogenes Stresserleben – wahrgenommene Bedrohung***

Das schulbezogene Bedrohungserleben nimmt in der Interventionsgruppe im Vergleich zur Kontrollgruppe, die keine bedeutsame Veränderung berichtet, über beide Projektjahre hinweg signifikant ab. Die diskontinuierlichen Veränderungsmodelle weisen darauf hin, dass dieser bedeutsame Gruppenunterschied wie auch schon bei der Prüfungsängstlichkeit vor allem auf die Entwicklung im ersten Projektjahr zurückzuführen ist, wo sich die wahrgenommene schulbezogene Bedrohung in der Interventionsgruppe verringert und in der Kontrollgruppe sogar leicht ansteigt (siehe auch Abbildung 4 im Anhang E). Im zweiten Jahr reduziert sich das Bedrohungserleben in beiden Gruppen – nach wie vor in der Interventionsgruppe stärker als in der Kontrollgruppe; der Gruppenunterschied in der Veränderung ist jedoch nicht signifikant. Dass sich sowohl bei der Prüfungsängstlichkeit als auch bei der wahrgenommenen Bedrohung signifikante, erwartungskonforme Gruppenunterschiede in der Veränderung bezogen auf das erste Projektjahr zeigen, war theoretisch zu erwarten. Denn dem Angst- und Stresserleben ist vor allem die Bedrohungskomponente gemeinsam (siehe Abschnitt 2.4.5). Prüfungsängstliche Schüler neigen dazu, Leistungsanforderungen als Bedrohung zu bewerten, sodass Leistungssituationen für diese Personen auch häufiger mit stressbezogenem situativem Bedrohungserleben einhergehen (Jerusalem, 1990). Die Reduktion der wahrgenommenen schulbezogenen Bedrohung im ersten Projektjahr geht, wie bei der Prüfungsängstlichkeit diskutiert, mit dem Einsatz der angst- und bedrohungsreduzierenden Module einher.

### ***Schulbezogenes Stresserleben – wahrgenommener Verlust***

Wie beim Herausforderungserleben sind auch beim Verlusterleben zunächst die signifikant unterschiedlichen Ausgangswerte zu beachten. Der wahrgenommene, schulbezogene Verlust wird zu Beginn des Projekts von den Interventionsschülern signifikant geringer eingeschätzt als von den Kontrollschülern. Ähnlich wie beim Herausforderungserleben könnte dieser Unterschied möglicherweise mit dem Wissen um die Teilnahme der Lehrkräfte an der FOSIS2-Intervention erklärt werden. Dieses Wissen könnte bei den Interventionsschülern dazu beigetragen haben, das schulbezogene Verlusterleben aufgrund der Chance, an einem Innovationsprojekt teilzunehmen, geringer zu bewerten als die Kontrollschüler. Die Ergebnisse zu den Gruppenunterschieden in der Entwicklung des wahrgenommenen Ver-

lusts müssen vor dem Hintergrund der unterschiedlichen Ausgangswerte mit Vorsicht interpretiert werden.

Hinsichtlich der Veränderung über die Zeit finden sich weder im Rahmen der linearen noch der diskontinuierlichen Analysen signifikante Gruppenunterschiede und die grafischen Darstellungen (siehe Abbildung 5 im Anhang E) belegen in der Interventions- und Kontrollgruppe einen nahezu identischen Verlauf des schulbezogenen Verlusterlebens, das im ersten Projektjahr zu- und im zweiten Projektjahr leicht abnimmt.<sup>58</sup> Diese Entwicklung entspricht dem beschriebenen Anstieg von Risikofaktoren für die Motivation nach dem Schulwechsel auf die weiterführende Schule (Wigfield et al., 2011; Wigfield et al., 2012). Anders als bei der Prüfungsängstlichkeit und dem Bedrohungserleben scheinen die FOSIS2-Maßnahmen der Zunahme im Verlusterleben nach dem Schulwechsel nicht entgegenzuwirken. Möglicherweise hängt dieser ausbleibende Effekt ebenfalls mit zu hohen, nicht erfüllten Erwartungen der Interventionsschüler an die FOSIS2-Lehrerfortbildungen zusammen, sodass der Anstieg des Verlusterlebens nicht verhindert werden konnte. In zukünftigen Studien sollten z. B. im Rahmen von explorativen, qualitativen Interviews die Schülererwartungen an Innovationsprojekte wie FOSIS2 mit erfasst werden und darüber hinaus der Zusammenhang zwischen den Wirkungen des FOSIS2-Programms und dem Verlusterleben, eventuell auch mithilfe modulspezifischer Evaluationsstudien, erneut überprüft werden.

### ***Binnendifferenzierung***

Bei der Interpretation der Ergebnisse zur wahrgenommenen Binnendifferenzierung im Unterricht ist die eingeschränkte psychometrische Qualität der Skala zu beachten. Nach Rost (2013) muss die Reliabilität bei Gruppenvergleichen im Rahmen von Feldforschungsprojekten zwar nicht unbedingt so hoch ausfallen wie z. B. bei individualdiagnostischen Studien. Sie sollte aber mindestens eine Ausprägung von  $\alpha = .55$  erreichen. Mit Werten von  $\alpha = .56$  zu T1,  $\alpha = .46$  zu T2 und  $\alpha = .54$  zu T3 kann die interne Konsistenz der Skala Binnendifferenzierung diese Forderung nicht ganz einhalten. Vor dem Hintergrund der geringen Reliabilität können die im Folgenden diskutierten Ergebnisse somit nur als erste Hinweise auf mögliche Wirkungszusammenhänge bewertet werden und sollten in künftigen Studien überprüft werden, nachdem die psychometrische Qualität der Skala verbessert wurde.

---

<sup>58</sup> Angesichts der sehr kleinen ICC von 0.022 (Hox, 2010) war dieser nicht vorhandene Effekt, wie bei der Selbstwirksamkeitserwartung (siehe Fußnote 57), bereits nach der Modellierung des Nullmodells abzusehen.

Im Verlauf der beiden Projektjahre weisen die linearen Wachstumskurvenmodelle in der Interventionsgruppe auf eine etwas stärkere Zunahme der wahrgenommenen Binnendifferenzierung hin als in der Kontrollgruppe. Der Gruppenunterschied ist jedoch nicht signifikant. Im Rahmen der diskontinuierlichen Modelle wird der Gruppenunterschied in der Veränderung bezogen auf das erste Projektjahr signifikant und die Aufnahme des Prädiktors ‚Gruppe‘ trägt zu einer Varianzaufklärung von 54 % bei: Die Interventionsschüler nehmen im ersten Jahr einen Anstieg der Binnendifferenzierung im Unterricht wahr, wohingegen die Kontrollschüler einen leichten Abfall berichten. Für das zweite Jahr zeigen sich keine bedeutsamen Gruppenunterschiede in der Veränderung und die Aufnahme des Gruppenprädiktors führt hier zu einer Varianzaufklärung von nur 6 %; die Gruppenunterschiede in der Entwicklung der wahrgenommenen Binnendifferenzierung sind im zweiten Projektjahr folglich als nicht relevant einzustufen. Wie bereits bei der Prüfungsängstlichkeit und der wahrgenommenen Bedrohung, wo sich ebenfalls nur für das erste Projektjahr signifikante Gruppenunterschiede in der Veränderung finden, lässt sich vermuten, dass der bedeutsame Zuwachs in der wahrgenommenen Binnendifferenzierung im ersten Jahr durch den Einsatz der spezifischen FOSIS2-Module in diesem Zeitraum begründet werden kann. Denn insbesondere das zweite und dritte Modul zu den Themen Transparenz und Lern- und Leistungsraum sollen die Binnendifferenzierung im Unterricht erhöhen: Durch die konkreten Angaben zu Anforderungen und Vorbereitungsmöglichkeiten im Transparenzpapier ist es dem Schüler leichter möglich, sich entsprechend seinen individuellen Fähigkeiten und seinem eigenen Lerntempo auf die Prüfung vorzubereiten. Auch der Lernraum soll den Schülern die Chance eröffnen, den eigenen Fähigkeiten angemessene Aufgaben auszuwählen und diese im eigenen Tempo zu bearbeiten, wobei dieser Zusammenhang durch die Analysen zum Implementationseinfluss in Frage gestellt wird. Diese belegen nämlich bei den Interventionsschülern einen negativen Zusammenhang zwischen der wahrgenommenen Umsetzung des Moduls Lern- und Leistungsraum und der Binnendifferenzierung, der im Abschnitt 8.2.2 diskutiert wird.

Im Gegensatz zu den anderen in der Arbeit untersuchten Variablen, bei denen es sich um internale, personale Ressourcen und Risikofaktoren handelt, ist die Binnendifferenzierung eine externe Ressource der Lernumwelt und kann ähnlich den FOSIS2-Maßnahmen auch als Förderstrategie für den Unterricht beschrieben werden. Dementsprechend wird das Konstrukt in Kapitel 2 auch unter dem Abschnitt 2.5 „Strategien und Ansätze zur Motivationsförderung in der Schule“ erläutert. Es läge nahe, die Binnendifferenzierung im Unterricht als Mediatorvariable zu untersuchen und deren Einfluss auf die Kriterien Selbstwirk-

samkeitserwartung, Prüfungsängstlichkeit und Stresserleben zu überprüfen. In der Arbeit wurde die Binnendifferenzierung jedoch in den Prä-, Zwischen- und Postbefragungen als Evaluationskriterium erhoben und es wurde angenommen, dass der Einsatz der FOSIS2-Maßnahmen nicht nur zu einer Veränderung personaler Variablen beiträgt, sondern auch die Lernumwelt selbst beeinflusst und somit z. B. zu einer Verbesserung von Unterrichtsstrategien wie der Binnendifferenzierung führt, die kein expliziter Bestandteil des FOSIS2-Programms sind. Bei künftigen Studien sollte nicht nur die Reliabilität der Skala verbessert, sondern die Binnendifferenzierung auch in den modulspezifischen Rückmeldungen erfasst und als Mediatorvariable überprüft werden.

Über alle Zielkriterien hinweg zeigen sich im Rahmen der diskontinuierlichen Mehrebenenanalysen nur für das erste Projektjahr signifikante, erwartungskonforme Gruppenunterschiede in der Veränderung. Es ist anzunehmen, dass diese bedeutsamen Unterschiede nicht nur mit den Inhalten der im ersten Projektjahr implementierten FOSIS2-Module zu den Themen Klassenklima, Transparenz und Lern- und Leistungsraum zusammenhängen, sondern auch mit dem Ausmaß der Umsetzung der Maßnahmen. Die deskriptiven Statistiken zur Teilnahme an den Workshops, zur Inanspruchnahme von Beratungssitzungen sowie zur von den Schülern wahrgenommenen Implementation weisen auf ein geringeres Umsetzungsverhalten der Lehrkräfte im zweiten Projektjahr hin. Das Ausmaß der Implementation und mögliche Zusammenhänge mit der Programmwirksamkeit werden im nachfolgenden Abschnitt diskutiert.

## **8.2 Diskussion der Ergebnisse zur Wirksamkeitsevaluation in Abhängigkeit von der wahrgenommenen Implementation**

Ein weiteres Ziel der vorliegenden Arbeit ist die Untersuchung der Programmwirksamkeit in Abhängigkeit von der von den Schülern wahrgenommenen modulspezifischen Umsetzung der FOSIS2-Maßnahmen mittels kovarianzanalytischer Veränderungsmodelle. Bevor die Ergebnisse dieser Analysen erläutert werden, sollen zunächst die deskriptiven Statistiken zur Implementation diskutiert werden.

### **8.2.1 Deskriptive Beschreibung der Implementation**

Eine kontinuierliche Programmimplementation erweist sich nach Payne (2009) insbesondere bei schulbasierten Präventionsprogrammen als schwierig. In einem Überblicksartikel zur Nachhaltigkeit des Umsetzungsverhaltens von Lehrkräften als Mediatoren beschreiben auch Han und Weiss (2005) eine rückläufige Implementation als typisches Problem von



schulbasierten Präventionsprogrammen, wobei ein intensives Training und ein angemessenes Feedback zur Umsetzung eine nachhaltige Innovationsbereitschaft der Lehrer fördern können. Im FOSIS2-Projekt sollten ein intensives Training sowie Möglichkeiten des Feedbacks und der Reflexion von Schwierigkeiten im Rahmen der regelmäßig stattfindenden Workshops, der Beratungssitzungen und der Unterrichtshospitationen durch die wissenschaftliche Begleitung realisiert werden. Auch das übersichtliche und leicht verständliche Begleitmaterial sollte die Durchhaltungsmotivation der Lehrkräfte stärken und somit zu einer kontinuierlichen Implementation der FOSIS2-Förderstrategien beitragen. Die deskriptive Auswertung der Implementation (Abschnitt 7.2.1) legt nahe, dass eine anhaltend hohe Umsetzung der FOSIS2-Förderstrategien nur teilweise gelang: Die Teilnahme der Lehrkräfte an den Workshops und auch die Anzahl der in Anspruch genommenen Beratungssitzungen sowie die von den Schülern wahrgenommene Umsetzung der FOSIS2-Förderstrategien fallen im zweiten Projektjahr geringer aus als im ersten (siehe Tabelle 18 bis Tabelle 20). Es ist somit anzunehmen, dass im zweiten Projektjahr das tatsächliche Implementationsverhalten der Lehrkräfte nachließ und weniger FOSIS2-Förderstrategien Eingang in den regulären Unterricht gefunden haben als im ersten Projektjahr.

Die im zweiten Jahr geringere Beteiligung der Lehrkräfte an den Workshops könnte im Sinne eines „Ermüdungseffekts“ mit einer eventuellen Überforderung durch das sukzessive Hinzukommen der FOSIS2-Module und die damit stetig wachsende Anzahl an neu erworbenen Unterrichtsstrategien zusammenhängen.<sup>59</sup>

Die Inanspruchnahme von modulspezifischen Beratungen (siehe Tabelle 19) hängt vermutlich zum einen mit der Komplexität der Modulinhalte und den damit verbundenen mehr oder weniger auftretenden Schwierigkeiten bei der Umsetzung zusammen. Zum anderen kann sie ebenfalls Ausdruck der Innovationsbereitschaft der am Projekt beteiligten Lehrkräfte sein. Über alle Schulen hinweg ist die Inanspruchnahme von Beratungstreffen im ersten Projektjahr höher als im zweiten. Dies stützt die Annahme, dass im zweiten Projektjahr, der geringeren Beteiligung an den Workshops entsprechend, weniger Lehrkräfte Förderstrategien in ihren Unterricht implementiert haben als im ersten. Am häufigsten ist die Inanspruchnahme von Beratungen beim ersten Modul zum Thema Klassenklima ( $n = 10$ ). Vermutlich hängt diese relativ hohe Anzahl an Beratungstreffen einerseits mit der zu Be-

---

<sup>59</sup> Bei der Bewertung der Beteiligungsraten ist zu beachten, dass die Schulleitung der Projektschule 3 zu Beginn des FOSIS2-Projekts das gesamte Kollegium und nicht nur die Zielgruppe der in den siebten Klassen unterrichtenden Lehrkräfte aufgefordert hatte, am ersten Workshop des FOSIS2-Projekts teilzunehmen. Erst ab dem zweiten Workshop beteiligten sich an dieser Schule nur die Lehrkräfte, die auch in den siebten Klassen unterrichteten. So erklärt sich der Rückgang der Beteiligung um 50 % an der Schule 3 (siehe Tabelle 18).

ginn des FOSIS2-Projekts noch großen Innovationsbereitschaft und Teilnahmemotivation der Lehrkräfte zusammen. Andererseits hatte das Thema „Klimaverbesserung“ möglicherweise auch gerade angesichts der Neukonstitution der siebten Klassen eine besondere Relevanz für die Lehrkräfte, die sich von den klimaförderlichen Maßnahmen vielleicht einen verbesserten Start der Schüler-Schüler- und Schüler-Lehrer-Beziehungen versprochen haben. Zum Modul Transparenzpapier fand nur eine Beratungssitzung in der Schule 3 statt. Die Maßnahme der Einführung eines Transparenzpapiers vor Klassenarbeiten erfordert von allen Förderstrategien des FOSIS2-Projekts den geringsten Zeitaufwand und kann nach genauen Vorgaben erarbeitet werden. Folglich sind bei der Umsetzung weniger Schwierigkeiten zu erwarten als bei anderen Modulen und der geringe Bedarf an Beratung und Hilfestellungen bei der Implementation des Transparenzpapiers ist somit nachvollziehbar. Für die Einführung des sogenannten Lernraumes in den Unterricht, eine Maßnahme, die etwas komplexer ist als das Transparenzpapier und stärker in die Unterrichtsroutinen eingreift, wurden über alle Schulen hinweg wieder etwas mehr Beratungssitzungen in Anspruch genommen. Beim vierten Modul zum Thema Selbstbestimmung, das die Einführung von Wahlmöglichkeiten in den Unterricht vorsieht und einen vergleichsweise geringen Änderungsbedarf im Unterrichtsverhalten der Lehrkräfte erfordert, wurde entsprechend der Simplizität der Maßnahme nur eine Beratungssitzung in der Schule 3 in Anspruch genommen. Obwohl insbesondere die letzten beiden Module zu den Themen Kooperatives Lernen und Selbstregulation als die komplexesten Module bewertet werden können, bei deren Umsetzung die meisten Schwierigkeiten zu erwarten waren, wurden für diese Module nur jeweils drei beziehungsweise zwei Beratungssitzungen in Anspruch genommen.

Die von den Schülern wahrgenommene modulspezifische Implementation der FOSIS2-Maßnahmen fällt ebenfalls im ersten Projektjahr größer aus als im zweiten (siehe Tabelle 20). Am höchsten ist die Wahrnehmung der Umsetzung der Förderstrategie Transparenzpapier und Lern- und Leistungsraum. Dies entspricht den Beobachtungen der wissenschaftlichen Begleitung in den Projektschulen. Denn nach Berichten der Lehrkräfte und auch der Schüler wurden die Transparenzpapiere so gut angenommen, dass sie teilweise auch von Schülern aus Klassen eingefordert wurden, die nicht am Projekt teilnahmen. In der Schule 3 wurden die Transparenzpapiere nach einiger Zeit sogar in allen Klassen implementiert. Teilweise erklärten sich Schüler bereit, die Transparenzpapiere in Absprache mit der unterrichtenden Lehrkraft selbst zu erstellen und für die Klasse zu vervielfältigen. Auch der Einsatz des Lernraums wurde von den Interventionsschülern deutlich wahrgenommen.

Dass die von den Schülern wahrgenommene Implementation bei den Modulen zu den Themen Transparenz und Lern- und Leistungsraum am höchsten ausfällt, passt zu den oben diskutierten Ergebnissen der diskontinuierlichen Veränderungsanalysen, wonach sich nur im ersten Projektjahr bedeutsame Effekte zeigen. Trotz des geringen Aufwands der Maßnahmen des vierten Moduls zur Förderung der Selbstbestimmung wurden diese von den Schülern am wenigsten wahrgenommen. Dies kann damit zusammenhängen, dass das Modul Selbstbestimmung zu Beginn des Schuljahres 2010/2011 mit der in Abschnitt 6.1.3 beschriebenen Schulreform zusammenfiel. Wie alle Berliner Oberschulen mussten sich die Projektschulen im Schuljahr 2010/2011 einer Schulreform unterziehen, bei der die Hauptschulen mit den Real- beziehungsweise Gesamtschulen zu sogenannten Integrierten Sekundarschulen zusammengelegt wurden (Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin, 2014). Zwar waren die Projektklassen des FOSIS2-Projekts nicht von der Schulreform betroffen, denn die bereits begonnenen Jahrgänge wurden nach dem alten Schulsystem fortgeführt. In den Beratungssitzungen berichteten die Lehrkräfte jedoch trotzdem immer wieder über mit der Schulreform zusammenhängende Ängste vor zu großen Klassen, mehr verhaltensauffälligen Schülern und einer damit verbundenen Überforderung. Auch wenn sich diese Befürchtungen nicht direkt auf die Projektklassen, sondern auf die neuen siebten Jahrgänge bezogen, könnten die Ängste Auswirkungen auf das Unterrichtsverhalten der Projektlehrer und somit auch auf die Umsetzung der FOSIS2-Förderstrategien gehabt haben. Möglicherweise waren die Lehrer angesichts der Schulreform und ihrer damit verbundenen Ängste zu Beginn des zweiten Projektjahres besonders restriktiv und weniger offen, den Schülern Wahlmöglichkeiten zuzugestehen. Generell lässt sich vermuten, dass Fortbildungen, die zu Beginn des Schuljahres stattfinden, aufgrund der mangelnden Routine (neuer Stunden- und Raumplan, neue Klassenzusammensetzung und neue Kollegen) weniger Eingang in den Unterricht finden als Fortbildungen, die etwas später im Schuljahr angesiedelt sind. Dies könnte auch erklären, warum die Förderstrategien zur Verbesserung des Klassenklimas zu Beginn des ersten Schulprojektjahres von den Schülern weniger stark wahrgenommen wurden als die etwas später im Jahr angesiedelten Module. Im Vergleich zum Modul Selbstbestimmung ist die wahrgenommene Umsetzung der letzten beiden Module Kooperatives Lernen und Selbstregulation wieder etwas höher, jedoch nicht so hoch wie die der Module Transparenz und Lern- und Leistungsraum.

### **8.2.2 Kovarianzanalytische Veränderungsmodelle**

Die Frage nach der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation wurde beispielhaft für das erste Projektjahr mittels kovarianzanalytischer Veränderungsmodelle untersucht. Zu beachten ist, dass die Implementationsdaten nur für die Interventionsschüler vorliegen. Bei den Analysen zum Einfluss der Implementation handelt es sich folglich nicht um Interventions- und Kontrollgruppenvergleiche und die Ergebnisse können somit nur als erste Hinweise auf mögliche Zusammenhänge zwischen dem Ausmaß der Implementation und der Programmwirksamkeit bewertet werden. Da die Implementation modulspezifisch erfasst wurde, werden die Ergebnisse in der Reihenfolge der Module des ersten Projektjahres diskutiert, sollten aber aufgrund der fehlenden Kontrollgruppe nicht als modulspezifische Wirkungen im Sinne einer Kausalität verstanden werden.

#### ***Klassenklima***

Hinsichtlich der wahrgenommenen Implementation des Moduls zur Förderung des Klassenklimas zeigt sich im Rahmen der kovarianzanalytischen Veränderungsmodelle ein signifikanter, positiver Zusammenhang mit dem Herausforderungserleben. Das heißt, unter Kontrolle der Ausgangswerte fühlen sich die Interventionsschüler mit wachsendem Implementationsausmaß zunehmend positiv durch die Schule herausgefordert. Dieser Zusammenhang könnte mit einem tatsächlich durch die Maßnahmen des Moduls verbesserten Klima begründet werden: Nach Saldern und Littig (1987) hängt das Klima von der Beschaffenheit der Schüler-Schüler- und der Lehrer-Schüler-Beziehung sowie von Merkmalen des Unterrichts ab. Es ist anzunehmen, dass Schüler, die eine hohe Zufriedenheit mit diesen drei Merkmalsbereichen aufweisen und sich durch ihre Mitschüler und Lehrkräfte unterstützt fühlen, auch eher positive Herausforderungen suchen. Neben diesem postulierten Wirkungsweg könnten die Schüler auch die Maßnahmen zur Klimaförderung (Zielscheibe zur Diagnose des Klassenklimas ausfüllen und auswerten, Klassenregeln und Konsequenzen bei Regelbruch erarbeiten, Verantwortung übernehmen; siehe Abschnitt 4.3.1) selbst als positive Herausforderungen wahrgenommen haben.

#### ***Transparenz***

Bezüglich der wahrgenommenen Umsetzung des Transparenzmoduls belegen die kovarianzanalytischen Veränderungsmodelle einen signifikanten, positiven Zusammenhang mit der Binnendifferenzierung. Unter Kontrolle der Ausgangswerte nehmen die Interventions-

schüler mit wachsendem Implementationsausmaß auch mehr Binnendifferenzierung im Unterricht wahr. Dies war insofern zu erwarten, als dass die Transparenzpapiere den Schülern eine gezielte, ihren eigenen Fähigkeiten und Leistungsniveaus entsprechend binnendifferenzierte und individualisierte Vorbereitung auf Klassenarbeiten ermöglichen sollen.

### ***Lern- und Leistungsraum***

Die Ergebnisse der kovarianzanalytischen Modelle zeigen einen positiven, signifikanten Zusammenhang auf zwischen der wahrgenommenen Implementation des Moduls zum Lern- und Leistungsraum und der schulbezogenen Herausforderung sowie der schulischen Selbstwirksamkeitserwartung. Das heißt, mit wachsendem Ausmaß an wahrgenommenen Maßnahmen zur Trennung von Lern- und Leistungsräumen geht bei den Interventionsschülern ein Zuwachs im Herausforderungserleben und der schulischen Selbstwirksamkeitserwartung einher. In den Lernräumen haben die Schüler die Gelegenheit zu üben, ohne dass sie benotet werden. Dadurch werden sie möglicherweise mutiger, sich Aufgaben zu suchen, die sie als positive Herausforderungen wahrnehmen. Wenn sie diese erfolgreich bewältigen, erfahren sie eigene Erfolge, sodass auch die schulische Selbstwirksamkeit gesteigert werden kann. Obwohl die Analysen der oben erläuterten diskontinuierlichen Veränderungsmodelle bei den Interventionsschülern, verglichen mit der Kontrollgruppe, im ersten Projektjahr erwartungskonform eine signifikant günstigere Entwicklung in der Binnendifferenzierung aufzeigen, findet sich bei den kovarianzanalytischen Modellen zur Implementation interessanterweise ein negativer, signifikanter Zusammenhang zwischen der wahrgenommenen Implementation des Moduls zum Lern- und Leistungsraum und der Binnendifferenzierung. Dieser negative Zusammenhang könnte darauf hinweisen, dass der Lernraum von den Lehrkräften des FOSIS2-Projekts nicht genutzt wurde, um differenzierte Aufgaben anzubieten. Die Checkitems zur Erfassung der Implementation des Lern- und Leistungsraumes fokussieren auf die Möglichkeit, im Unterricht lernen und üben zu können, ohne dass dafür Noten vergeben werden (siehe Anhang B5). Die Items erfragen nicht, ob dabei beispielsweise unterschiedliche Lerntempi beachtet wurden oder differenzierte Aufgaben zur Verfügung standen. Möglich wäre, dass Schüler mit zunehmend mehr Gelegenheiten, im Unterricht zu üben und zu lernen, ohne benotet zu werden, bewusster wahrnehmen, ob dabei ihre individuellen Fähigkeiten und Lerntempi berücksichtigt werden, und somit das Ausbleiben oder die Verringerung differenzierter Maßnahmen im Unterricht deutlicher erleben.

Zwischen der wahrgenommenen Implementation der Module des ersten Projektjahres und der Prüfungsängstlichkeit sowie dem Bedrohungs- und Verlusterleben finden sich keine signifikanten Zusammenhänge. Dies irritiert insofern, als dass sich im Rahmen der diskontinuierlichen Analysen bei den Interventionsschülern, verglichen mit der Kontrollgruppe, über das erste Projektjahr hinweg gerade in der Prüfungsängstlichkeit und der wahrgenommenen Bedrohung signifikant günstigere Entwicklungen zeigen. Eine mögliche Begründung für die ausbleibenden Zusammenhänge zwischen dem Ausmaß der wahrgenommenen Implementation und der Prüfungsängstlichkeit sowie der wahrgenommenen Bedrohung könnte sein, dass hier schon eine geringe Umsetzung der Maßnahmen ausreicht, um Effekte zu erzielen und ab einem bestimmten Schwellenwert eine Variation in der Implementation keinen zusätzlichen Gewinn bringt. Möglicherweise ist der Zusammenhang zwischen dem Ausmaß der Umsetzung und der Programmwirksamkeit nicht zwangsläufig linear und ein „Mehr“ an Implementation hat nicht immer einen positiven Einfluss. In zukünftigen Studien sollten daher auch nicht lineare Zusammenhänge z. B. wiederum mittels diskontinuierlicher Mehrebenenanalysen möglichst auch im Rahmen von Interventions- und Kontrollgruppenvergleichen überprüft werden.

### **8.3 Bemerkungen zur Durchführung des FOSIS2-Projekts**

Die FOSIS2-Maßnahmen, die zur Förderung des Klassenklimas, der Transparenz und zur Erhöhung von Lernräumen im Unterricht sowie zu einem verstärkten Einsatz selbstbestimmter, kooperativer und selbstregulierter Lernformen beitragen sollen, stehen aktuellen politischen Bestrebungen für die Schule teilweise entgegen. So werden politisch mehr soziale Vergleiche (z. B. PISA) und eine noch häufigere, standardisierte Erfassung von schulischen Leistungen gefordert, wodurch der Leistungsdruck auf Schüler- und Lehrerschaft zunimmt (Wigfield et al., 2012). Dadurch ist in der Lehrerschaft nicht immer die Offenheit für Fortbildungsprogramme wie FOSIS2 gegeben, was die Projektdurchführung erschweren kann. Auf der Basis von Erfahrungsberichten der wissenschaftlichen Begleitung soll deshalb an dieser Stelle dargestellt werden, welche Eigenschaften des FOSIS2-Projekts die Projektdurchführung erleichtert haben und welche Aspekte verbesserungswürdig scheinen. Zu beachten ist, dass es sich bei diesen Schilderungen nicht um systematisch evaluierte Bestandteile des FOSIS2-Projekts handelt, die Bemerkungen geben vielmehr den subjektiven Eindruck der Projektmitarbeiter wieder. Eine systematische Evaluation der hier als erfolgreich beziehungsweise verbesserungswürdig empfundenen Merkmale der Projektdurchführung steht aus.

Die im Vergleich zu den Vorgängerprojekten veränderten Rahmenbedingungen haben sich aus Sicht der wissenschaftlichen Begleitung des FOSIS2-Projekts weitestgehend als positiv erwiesen. So scheint die veränderte zeitliche und organisatorische Struktur der Interventionsinhalte von Vorteil zu sein: Anders als bei vorherigen Studien, wo den Lehrkräften in umfassenden Fortbildungsveranstaltungen eine Vielzahl verschiedener Förderstrategien vermittelt wurde, besteht das FOSIS2-Programm aus sechs kompakten, inhaltlich in sich kohärenten Modulen. Für die kurzen, im Abstand von circa drei Monaten stattfindenden Modulworkshops ließen sich leicht Termine finden und durch die sukzessive Einführung der Maßnahmen konnten diese dem Eindruck nach leichter in den Unterrichtsalltag integriert werden. Mithilfe des umfangreichen Unterstützungs- und Beratungsangebots, welches für jede Interventionsschule einen Projektbetreuer vorsah, der im Rahmen von Einzel-, Gruppen-, E-Mail- und Telefonberatungen für Fragen zu den FOSIS2-Inhalten zur Verfügung stand, konnten Schwierigkeiten bei der Umsetzung schnell geklärt werden. Die Unterstützung seitens der wissenschaftlichen Begleitung wurde von den Lehrkräften zwar unterschiedlich stark angenommen. Die Projektlehrer, die kontinuierlich an den Workshops und den Beratungsmöglichkeiten teilnahmen, scheinen die FOSIS2-Maßnahmen jedoch vermehrt umgesetzt und internalisiert zu haben. Auch die regelmäßige Erfassung der Implementation mittels modulspezifischer, kurzer, standardisierter Fragebögen scheint im Vergleich zu den in Vorgängerprojekten eingesetzten Logbüchern eine günstige Variante zur Erfassung des Implementationsgeschehens zu sein. Zumindest konnten dadurch höhere Rücklaufquoten erzielt werden. Trotz der bewussten Rekrutierung von für Innovationen wie FOSIS2 offenen Schulleitungen wird die Unterstützung durch die Schulleitung von den Projektmitarbeitern als verbesserungswürdig bewertet. So haben die Lehrkräfte über den gesamten Projektzeitraum immer wieder eine zu geringe Unterstützung durch die Schulleitung, insbesondere einen mangelnden Ausgleich für die Teilnahme am FOSIS2-Projekt und den damit verbundenen zusätzlichen Zeitaufwand beklagt. Für die Teilnahme an langfristigen Fortbildungsmaßnahmen, wie dem FOSIS2-Programm, sollten künftig Freistunden zur Verfügung gestellt werden und auch die Treffen von klassenübergreifenden Lehrerteams sollten möglichst in den regulären Stundenplan integriert werden. So würde das besondere Engagement von Lehrkräften, die an Fortbildungen teilnehmen und sich offen für die Kooperation im Team zeigen, belohnt und nicht durch einen zusätzlichen Zeitaufwand erschwert.

## 8.4 Einschränkungen der Studie

Nach Patry und Hager (2000) liegt ein zentrales Dilemma von wissenschaftlichen Evaluationsstudien in der Schwierigkeit, die in Kapitel 3 geforderten methodischen Forderungen einzuhalten und gleichzeitig den Bedürfnissen der Praktiker, die das zu evaluierende Programm umsetzen sollen, gerecht zu werden. Die Praktiker erhoffen sich durch Evaluationen in der Regel einfache Antworten und Lösungen für teilweise sehr komplexe in der Praxis auftretende Fragen und Probleme. Zudem wünschen sie sich Interventionen und Evaluationen, die sich mit geringem Aufwand in ihren praktischen Alltag integrieren lassen. Diese Bedürfnisse können unter Einhaltung der geschilderten Forderungen an Evaluationsstudien, die aufwendige Studiendesigns und die Einführung von Vergleichsgruppen und Kontrollbedingungen beinhalten, oft nicht erfüllt werden. Denn „je stärker man versucht, sich wissenschaftlichen Kriterien und damit vor allem der möglichst eindeutigen Interpretierbarkeit anzunähern, desto aufwendiger dürften empirische Studien in der Regel werden“ (Patry & Hager, 2000, S. 267). Das FOSIS2-Projekt stellt einen Versuch dar, einen Kompromiss zwischen der Wissenschaftlichkeit der Untersuchungsanlage und der Durchführbarkeit zu schließen, der zugunsten der Durchführbarkeit gewisse methodische Einschränkungen nach sich zieht. Diese durch das Design, die Stichprobe und die Erfassung der Implementation bedingten Einschränkungen, die die Validität der Ergebnisse beeinflussen, werden nachfolgend erläutert.

### *Durch das Design bedingte Einschränkungen*

Beim FOSIS2-Projekt handelt es sich um ein Quasi-Experiment, bei dem die Zuteilung zur Interventions- und Kontrollgruppe nicht per Randomisierung erfolgen konnte. Nach Bortz und Döring (2006) sind quasi-experimentelle Designs typisch für Evaluationsstudien im Kontext Schule, wo die Schulklassen und Schulen natürliche Gruppen bilden und eine Randomisierung auch aus ethischen Gesichtspunkten oft nicht möglich ist, da die Schüler und Lehrer nicht zur Teilnahme an einer Intervention gezwungen werden können. Durch die fehlende Randomisierung und auch durch das Ausbleiben der von Hager (2008) geforderten behandelten Kontrollgruppe<sup>60</sup> kann die Kontrolle von Störfaktoren in der vorliegen-

---

<sup>60</sup> Im Rahmen von sogenannten isolierten Wirksamkeitsevaluationen, die wie die vorliegende Arbeit das Ziel verfolgen, die Effekte eines einzelnen Programms zu untersuchen, fordert Hager (2008) die Einführung einer behandelten Kontrollgruppe, die eine Intervention mit anderen Zielen, aber gleichen Randbedingungen als das zu evaluierende Programm erhalten soll. Nur so lassen sich die maßnahmenspezifischen Nettowirkungen des Zielprogramms ermitteln (Hager, 2008; siehe auch Abschnitt 3.1.6). Dieser Empfehlung konnte im Rahmen des FOSIS2-Projekts nicht nachgekommen werden, da die Einführung behandelter Kontrollgruppen aufgrund der Kosten und des Aufwands nicht umsetzbar war.



den Untersuchung nicht optimal gewährleistet werden. Störfaktoren sind nach Hager und Hasselhorn (2000) alle Prozesse, die zu einer Schwächung der intendierten Programmwirkung, also der internen Validität, beitragen können, z. B. *Reifungsprozesse*, *Selektionseffekte*, *Drop-Out-Prozesse*, *externe Einflüsse* oder auch der *Hawthorne Effekt*. Die genannten Störfaktoren sollen nachfolgend in Bezug zur vorliegenden Untersuchung diskutiert werden.

Mögliche Reifungsprozesse, das heißt eine durch die Entwicklung bedingte Reifung der Interventionsschüler, können als Störfaktoren insofern ausgeschlossen werden, als dass die Kontrollschüler im gleichen Alter über den Projektzeitraum denselben Entwicklungsprozess durchlaufen haben.

Im Gegensatz dazu scheinen Selektionseffekte im FOSIS2-Projekt nicht unwahrscheinlich: Aufgrund der Erfahrung aus Vorgängerprojekten wurden bewusst solche Schulen als Interventionsschulen rekrutiert, die sich besonders innovationsfreudig zeigten, ein eigenes Interesse an dem FOSIS2-Programm bekundeten und die Programmziele mit der eigenen Schulkultur verbinden konnten. Zwar wurden die Schulleiter gebeten, für die Kontrollgruppe Schulen zu benennen, die der jeweiligen Interventionsschule in Größe, Zusammensetzung der Schüler- und Lehrerschaft, in Schulkultur und anderen strukturellen Merkmalen ähneln. Diese Ähnlichkeiten wurden jedoch nicht systematisch kontrolliert.

Bezüglich der Beteiligung der Lehrkräfte an den Workshops sind darüber hinaus Drop-Out-Prozesse festzustellen, für die bereits als mögliche Erklärung eine abnehmende Innovationsbereitschaft diskutiert wurde (Abschnitt 8.2.1). Patry und Hager (2000) beschreiben eine mangelnde Innovationsbereitschaft von Praktikern, die dem Wunsch nach Veränderung und Neuerung der Wissenschaftler entgegensteht, als ein typisches Problem von Evaluationsstudien. Im FOSIS2-Projekt konnte in jeder Schule zumindest ein kleines Team an sehr aktiven, innovationsbereiten Lehrkräften etabliert werden, das kontinuierlich, mit Ausdauer und hoher Motivation über den gesamten Projektzeitraum am FOSIS2-Programm teilnahm.

Ein potentieller externer Störfaktor der vorliegenden Studie ist die Schulreform, deren mögliche Auswirkungen auf die Programmumsetzung zu Beginn des zweiten Schuljahres bereits beschrieben wurden (siehe Abschnitt 8.2.1).

Nicht auszuschließen ist außerdem, dass allein das Wissen der Interventionsschüler um die Teilnahme ihrer Lehrkräfte an den FOSIS2-Maßnahmen und nicht die eigentliche Umsetzung der FOSIS2-Förderstrategien in den Unterricht eine Veränderung der Zielkriterien

bewirkt haben könnte. Ein solcher typischerweise auch als Hawthorne Effekt bezeichneter Störfaktor wird auch von Martin (2008) in seiner Studie zur Evaluation eines Motivationsförderprogramms diskutiert: „It may be that a Hawthorne Effect can account for part of the gains observed. That is, simply being part of any new program [...] prompted students [...] perhaps to inflate their selfreports of motivation and engagement“ (S. 261). Nicht auszuschließen ist zudem, dass die Kontrollschulen im Projektzeitraum von sich aus an Lehrerfortbildungen teilnahmen, die möglicherweise sogar ähnliche Ziele wie das FOSIS2-Programm verfolgten und somit ebenfalls zu günstigen Veränderungen der hier untersuchten Schülervariablen beigetragen haben könnten. Angesichts der Länge des Projekts (zwei Schuljahre) muss sogar davon ausgegangen werden, dass zumindest vereinzelt auch die Kontrolllehrer an Lehrerfortbildungen teilgenommen haben. Dies wurde im FOSIS2-Projekt jedoch nicht kontrolliert und auch die modulspezifischen Rückmeldungen zur Erfassung der Umsetzung der FOSIS2-Maßnahmen wurden nur in den Interventionsschulen eingesetzt, weshalb für die Analysen zur Implementation keine Kontrollgruppe vorliegt.

Angesichts der erläuterten potentiellen Störfaktoren wären zur Verbesserung der internen Validität im Rahmen zukünftiger Studien eine behandelte Kontrollgruppe und eine sogenannte Cluster-Randomisierung wünschenswert, bei der die Zuteilung der Klassen zur Interventions- und Kontrollgruppe per Zufall erfolgt (Wirtz, 2014). Außerdem sollten bei künftigen Studien zur Evaluation der FOSIS2-Interventionen mögliche Fortbildungsmaßnahmen und deren Inhalte an den Kontrollschulen mit erfasst werden. Im besten Fall können Unterrichtsbeobachtungen in Interventions- und Kontrollschulen stattfinden, die eine genauere Analyse des Unterrichtsgeschehens und der Umsetzung von Interventionen ermöglichen (Blömeke, 2004; Petermann, 2014b). Diese erfordern jedoch eine hohe finanzielle und personelle Projektausstattung (Petermann, 2014b).

### ***Durch die Stichprobe bedingte Einschränkungen***

Eine weitere Einschränkung der vorliegenden Studie, die eine Verringerung der externen Validität nach sich zieht, stellt die relativ kleine Stichprobengröße dar. Zielgruppe der vorliegenden Untersuchung sind Schüler der siebten und achten Jahrgänge von sechs Berliner Oberschulen. Nach Hox (2010) sollten mindestens  $n = 30$  Einheiten auf dem höchsten Level vorliegen. Angesichts dieser Forderung und der geringen Anzahl an untersuchten Schulen ( $n = 6$ ) wird deutlich, warum die Schulebene in der vorliegenden Arbeit nicht als höchste Ebene berücksichtigt werden konnte. Dies schwächt die Aussagekraft der Ergebnisse insofern, als dass durch die Zugehörigkeit zu einer Schule bedingte Abhängigkeiten

nicht berücksichtigt werden konnten. Aus ökonomischer Perspektive haben bereits Cohen (1998) sowie Snijders und Bosker (1993) darauf hingewiesen, dass die genannten Anforderungen an die Anzahl der Einheiten pro Ebene im Hinblick auf die Kosten bei der Stichprobenziehung in der Praxis oft nicht einzuhalten sind. Dementsprechend lassen sich Studien finden, die Mehrebenenanalysen mit weniger als 30 Einheiten auf der höchsten Gruppenebene durchführen (z. B. Gelman, 2006; Magadi & Desta, 2011; Souvignier & Mokhlesgerami, 2005). Die höchste Ebene der in der vorliegenden Arbeit aufgestellten Modelle ist die Klassenebene mit 22 beziehungsweise bei den kovarianzanalytischen Modellen nur 10 Einheiten. Die Anzahl der Klassen konnte im vorliegenden Projekt aufgrund der begrenzten finanziellen und personellen Ressourcen nicht erhöht werden. Um aber für die Berechnungen zumindest die Stichprobengröße auf dem Schüler-Level zu maximieren, wurden für die linearen und diskontinuierlichen Modelle zur Evaluation der Wirksamkeit des FOSIS2-Programms im Rahmen sogenannter modellbasierter Maximum-Likelihood-Verfahren zum Umgang mit fehlenden Werten alle verfügbaren Schüler-Daten verwendet (siehe Abschnitt 6.3.2). Für die Beantwortung der Frage nach der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation wurde der fallweise Ausschluss eingesetzt und bei den Analysen nur Fälle mit vollständigen Werten eingeschlossen. Dadurch verringert sich die Schülerstichprobe. Mithilfe des Ansatzes der multiplen Imputation (siehe hierfür auch Lüdtke et al., 2007) hätte diese Reduktion der Stichprobe vermieden werden können. Die hier durchgeführten Analysen stellen jedoch lediglich einen ersten explorativen Versuch dar, die wahrgenommene modulspezifische Implementation seitens der Interventionsschüler zu untersuchen. Folglich werden die Ergebnisse der Implementationsanalysen als Pilot-Ergebnisse gewertet, die erstmalig Hinweise auf mögliche Wirkungszusammenhänge zwischen der modulspezifischen Implementationswahrnehmung der Interventionsschüler und den Kriterien liefern können. Aufgrund der explorativen Beschaffenheit der Analysen wurde in der vorliegenden Untersuchung auf die multiple Imputation verzichtet und der fallweise Ausschluss praktiziert. Da sich keine systematischen Zusammenhänge zwischen den fehlenden Werten und anderen erhobenen Variablen zeigen, scheint die Voraussetzung für den fallweisen Ausschluss, die MCAR-Annahme, haltbar zu sein (siehe Abschnitt 6.3.2).

Angeichts der für Mehrebenenanalysen kleinen Stichprobe sowie der nicht berücksichtigten Schulebene können die Ergebnisse dieser Arbeit lediglich als erste Hinweise auf mögliche Effekte des FOSIS2-Programms bewertet werden und sollten unbedingt an größeren, möglichst repräsentativen Stichproben überprüft werden. Insbesondere der Zusammenhang

zwischen der Implementation und der Programmwirksamkeit bedarf weiterer Untersuchungen im Rahmen größer angelegter Studien.

### ***Durch die Erfassung der Implementation bedingte Einschränkungen***

Patry und Hager (2000) weisen darauf hin, dass Unterrichtssituationen nur schwer zu standardisieren sind. Auch wenn, wie z. B. für das FOSIS2-Programm, bestimmte Manuale oder Handreichungen vorliegen, die eine einheitliche Einführung der Interventionen gewährleisten sollen, variiert in der Regel die praktische Umsetzung der Fördermaßnahmen zwischen den Lehrkräften (Durlak, 2013; Petermann, 2014b). In der vorliegenden Arbeit wurde diese Variation berücksichtigt, indem das Ausmaß durch die wahrgenommene Implementation seitens der Schüler erfasst und in die Wirksamkeitsanalysen einbezogen wurden. Nicht systematisch berücksichtigt wurde jedoch die Qualität der Umsetzung der FOSIS2-Fortbildungsinhalte. Eine „inkorrekte“ oder zumindest von den FOSIS2-Handreichungen abweichende Umsetzung, die ebenfalls zu einer Verringerung der Programmwirkung beigetragen haben kann, lässt sich dadurch nicht ausschließen. Wünschenswert wäre, dass in künftigen Studien auch die Qualität sowie eventuell auch weitere Kennwerte der Implementation (z. B. die Programmakzeptanz oder auch die Reichweite des Programms; siehe Abschnitt 3.2.2) möglichst unter Berücksichtigung unterschiedlicher Perspektiven (Selbstbericht der Lehrkräfte, Wahrnehmung der Schüler, Unterrichtsbeobachtung durch externe Personen) untersucht werden.

## **8.5 Zusammenfassung und Ausblick**

*„Much of the research within the field of achievement motivation is correlational or conducted in lab-like experimental contexts that do not reflect typical learning environments. Educational researchers need to collaborate with teachers and administrators to implement the suggested changes to the context and evaluate the effectiveness of these changes. Although this is certainly a large task and, [...] does not always provide the type of empirical data one may desire, it is essential that these theories of motivation are directly applied to students' learning“* (Linnenbrink & Pintrich, 2002, S. 323 f.).

Die vorliegende Studie greift die Forderung im vorangestellten Zitat auf und stellt einen Versuch dar, die Wirksamkeit theoriebasierter Maßnahmen zur Förderung ausgewählter motivational relevanter Konstrukte in der schulischen Praxis zu überprüfen. Nachfolgend werden die Ergebnisse der Studie zusammenfassend referiert und mögliche Anschlussfragen aufgezeigt.

Im Rahmen der linearen und diskontinuierlichen Analysen zur Evaluation der Wirksamkeit des FOSIS2-Programms zeigen sich im ersten Projektjahr bei den Interventionsschülern, verglichen mit den Kontrollschülern, eine signifikante Reduktion von Prüfungsängstlichkeit und Bedrohungserleben sowie ein bedeutsamer Anstieg in der wahrgenommenen Binnendifferenzierung. Somit lässt sich eine vor allem ängstlichkeits- und bedrohungsreduzierende Wirkung der FOSIS2-Maßnahmen im ersten Projektjahr festhalten. Der Effekt auf die Binnendifferenzierung sollte aufgrund der geringen Reliabilität der Skala unbedingt in künftigen Studien überprüft werden. Darüber hinaus erscheint es sinnvoll, die Binnendifferenzierung auch als Mediatorvariable zu testen (siehe Abschnitt 8.1). Obwohl sich bei der schulischen Selbstwirksamkeitserwartung und den Stress-Subskalen „schulbezogene Herausforderung“ und „Verlust“ ohne Berücksichtigung der Implementation keine signifikanten Effekte zeigen, können die erwartungskonformen Ergebnisse bei der Prüfungsängstlichkeit und der wahrgenommenen Bedrohung positiv beurteilt werden. Nach Kirkpatrick (1998) zeigen sich Veränderungen auf Schülerseite durch Lehrerfortbildungen nämlich häufig erst an letzter Stelle im Wirkungsverlauf. Insofern ist die angst- und bedrohungsreduzierende Wirkung bereits als ein erster Erfolg des FOSIS2-Programms zu bewerten. Weitere, größer angelegte Studien zur Evaluation des FOSIS2-Programms sind notwendig, um diese Wirkung zu überprüfen und weitere Effekte des Programms sowie modulspezifische Wirkungen zu ermitteln. Dabei sollte unbedingt auch die Implementation der Maßnahmen Berücksichtigung finden. Denn die Ergebnisse der Modelle zur Analyse der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation legen nahe, dass im Rahmen quasi-experimenteller Interventionsstudien nicht von einer einheitlichen Implementation der Interventionen durch die in der Schule tätigen Multiplikatoren ausgegangen werden kann und sich bedeutsame Veränderungen in der Interventionsgruppe in Abhängigkeit vom Ausmaß der wahrgenommenen Umsetzung zeigen. Auch wenn diese Ergebnisse aufgrund der fehlenden Kontrollgruppe und der kleinen Stichprobe nur als erste Hinweise auf einen Zusammenhang zwischen der Wirkung der FOSIS2-Maßnahmen und dem Implementationsausmaß bewertet werden können, stützen die Ergebnisse den Aufruf der Implementationsforschung, dass die Erfassung der Implementation im Rahmen von Evaluationsstudien unbedingt notwendig ist (Durlak, 2013; Durlak & DuPre, 2008; Gollwitzer & Jäger, 2009; Weissberg et al., 2003).

In methodischer Hinsicht belegen die Ergebnisse, dass die diskontinuierlichen Modelle den linearen Modellen hinsichtlich ihrer Vorhersagefähigkeit in der vorliegenden Studie überlegen sind. Die diskontinuierlichen Modelle verfügen über den besten Modellfit (der AIC

der diskontinuierlichen Veränderungs- und Evaluationsmodelle ist fast immer geringer als der AIC der linearen Modelle). Die Ergebnisse weisen somit darauf hin, dass die sehr häufig verwendete Modellierung von Veränderung mittels linearer Wachstumskurvenmodelle (Luhmann & Eid, 2013) im Rahmen ähnlich angelegter Längsschnittstudien nicht zwangsläufig die Modellierung der Wahl darstellt. Insbesondere wenn nicht-lineare Veränderungen vorliegen, keine Hypothesen über spezifische, höhere Zeitterme (z. B. quadratische oder kubische) existieren und mehrere Messzeitpunkte vorhanden sind, scheinen diskontinuierliche Modelle vielversprechend zu sein und sollten auch im Rahmen zukünftiger Studien zur Evaluation der FOSIS2-Maßnahmen zur Anwendung kommen.

Dabei könnten nicht nur die hier untersuchten Fragestellungen aufgegriffen, sondern auch Anschlussfragen untersucht werden. So wird beispielsweise derzeit in einer anderen im FOSIS2-Projekt angesiedelten Forschungsarbeit untersucht, welche Effekte das FOSIS2-Programm auf das Lehrerhandeln und möglicherweise auch auf personale Ressourcen und Risikofaktoren der Interventionslehrer hat. Von Interesse wäre neben der Untersuchung der Quantität der Umsetzung auch die Evaluation der Wirksamkeit in Abhängigkeit von der Qualität und möglichen anderen Kennwerten der Implementation mittels verschiedener Datenquellen. Inwieweit die von den Lehrkräften berichtete, die beobachtete und die durch die Schüler wahrgenommene Implementation korrelieren, wäre hierbei ebenfalls eine relevante Forschungsfrage. Auch wäre es wünschenswert zu überprüfen, ob die Auswirkungen der FOSIS2-Interventionen in Abhängigkeit vom Geschlecht, der Leistung oder anderer personaler Voraussetzungen (z. B. hohe oder niedrige Selbstwirksamkeit in den Ausgangswerten) differieren. Im Wirkmodell (Kapitel 5) wurde zudem aufgezeigt, dass die FOSIS2-Maßnahmen bei den Interventionsschülern nicht nur zu einer günstigen Entwicklung der hier untersuchten motivationalen Ressourcen und Risikofaktoren, sondern auch zu einer Verbesserung sozialer Variablen beitragen sollen. Inwieweit diese sozialen Programmziele erfüllt werden können, bleibt offen und muss noch überprüft werden.

Abschließend lässt sich festhalten, dass die Ergebnisse der vorliegenden Untersuchung vor allem eine ängstlichkeits- und bedrohungsreduzierende Wirkung der FOSIS2-Module des ersten Projektjahres belegen. Darüber hinaus weisen die Ergebnisse auch auf positive Zusammenhänge zwischen dem Ausmaß der von den Schülern wahrgenommenen Umsetzung der Module des ersten Projektjahres und den hier untersuchten Ressourcen für die schulische Motivation hin (z. B. bei der schulischen Selbstwirksamkeitserwartung und dem Herausforderungserleben). Die Kombination der Module des ersten Projektjahres zu den The-

men Klassenklima, Transparenz und Lern- und Leistungsraum scheint daher für eine Adaptation in anderen Schulen geeignet zu sein, wobei die deskriptive Auswertung der Daten zur von den Schülern wahrgenommenen Umsetzung nahelegt, dass insbesondere die Module Transparenz und Lern- und Leistungsraum implementiert wurden. Inwieweit sich eine ängstlichkeits- und bedrohungsreduzierende Wirkung sowie eventuell weitere Effekte auch auf die Einführung jeweils einzelner Module zurechnen lassen, muss im Rahmen modulspezifischer Evaluationsstudien noch überprüft werden. Hinsichtlich der Durchführung des FOSIS2-Projekts haben sich im Vergleich zu Vorgängerprojekten die hier gewählte Zusammenstellung der Maßnahmen in inhaltlich kohärente, kompakte Module, die den Lehrkräften im Rahmen von ein- bis dreistündigen Workshops vermittelt werden können, und insbesondere die enge Begleitung der Lehrkräfte durch die Projektmitarbeiter im Rahmen von Beratungen und Hospitationen bewährt und sollten auch bei künftigen Fortbildungsveranstaltungen und der Umsetzung der Maßnahmen in den Unterricht beibehalten werden.

Es bleibt zu hoffen, dass die beteiligten Lehrkräfte und Schüler der Interventionsschulen auch heute noch von den FOSIS2-Maßnahmen profitieren und als Multiplikatoren dazu beitragen, dass insbesondere die hier für die Prüfungsängstlichkeit und das Bedrohungserleben nachweislich bedeutsamen Module des ersten Projektjahres nach und nach auch von anderen Klassen der Projektschulen übernommen werden.

## Literaturverzeichnis

- Abel, S. (2013, 23. September). *Stressfaktor Schule. Drill oder Kuschelpädagogik?*. zdf-info-Kanal.
- Acee, T. W. & Weinstein, C. E. (2010). Effects of a Value-Reappraisal Intervention on Statistics Students' Motivation and Performance. *The Journal of Experimental Education*, 78 (4), 487–512.
- The American Institute of Stress. (2013). *What ist Stress?* Zugriff am 23.11.2013. Verfügbar unter <http://www.stress.org/what-is-stress>
- Ames, C. (1992a). Achievement goals and the classroom motivational climate. In D. H. Schunk & J. L. Meece (Hrsg.), *Student perceptions in the classroom* (S. 327–348). Hillsdale, New Jersey: Lawrence Erlbaum.
- Ames, C. (1992b). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261–271.
- Antonovsky, A. (1993). The structure and properties of the sense of coherence scale. *Social Science & Medicine*, 36 (6), 725–733.
- Aronson, E. (2014). *Jigsaw Classroom*. Zugriff am 15.01.2014. Verfügbar unter <http://www.jigsaw.org>
- Aronson, E., Blaney, N., Stephin, C., Sikes, J. & Snapp, M. (1978). *The Jigsaw classroom*. Beverly Hills: Sage Publications.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359–372.
- Atkinson, J. W. (1958). *Motives in fantasy, action, and society*. Princeton, NJ: Van Nostrand.
- Augstein, R. (Hrsg.). (2013, 22. April). Generation Stress [Themenheft]. *Der Spiegel online*, 17.
- Backhaus, O., Petermann, F. & Hampel, P. (2010). Effekte des Anti-Stress-Trainings in der Grundschule. *Kindheit und Entwicklung*, 19 (2), 119–128.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37 (2), 122–147.
- Bandura, A. (1986). *Social foundations of thought and action*. New Jersey: Prentice-Hall.
- Bandura, A. (1988). Self efficacy conception of anxiety. *Anxiety Research. An International Journal*, 1, 77–98.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology*, 52 (1), 1–26.
- Bauer, K.-O. (2007). Theorie und Methodologie der Evaluation an Schulen. In K.-O. Bauer (Hrsg.), *Evaluation an Schulen. Theoretischer Rahmen und Beispiele guter Evaluationsspraxis* (S. 13–51). Weinheim: Juventa-Verl.



- Baumann, T., Schneider, C., Vollmar, M. & Wolters, M. (2012). *Schulen auf einen Blick*. Wiesbaden: Statistisches Bundesamt.
- Beelmann, A. & Karing, C. (2014). Implementationsfaktoren und -prozesse in der Präventionsforschung: Strategien, Probleme, Ergebnisse, Perspektiven. *Psychologische Rundschau*, 65 (3), 129–139.
- Beyer, A. & Lohaus, A. (2006). *Stressbewältigung im Jugendalter. Ein Trainingsprogramm*. Göttingen: Hogrefe.
- Beyer, A. & Lohaus, A. (2007). Konzepte zur Stressentstehung und Stressbewältigung im Kindes- und Jugendalter. In I. Seiffge-Krenke & A. Lohaus (Hrsg.), *Stress und Stressbewältigung im Kindes- und Jugendalter* (S. 11–27). Göttingen: Hogrefe.
- Billings, A. G. & Moos, R. H. (1982). The role of coping responses and social resources in attenuating the stress of life events. *Journal of Behavioral Medicine*, 4, 139–157.
- Bilz, L., Ottava, V. & Ravens-Sieberer, U. (2013). Psychische Auffälligkeiten bei Schülerinnen und Schülern: Prävention und Früherkennung. In P. Kolip, A. Klocke, W. Melzer & U. Ravens-Sieberer (Hrsg.), *Gesundheit und Gesundheitsverhalten im Geschlechtervergleich. Ergebnisse des WHO-Jugendgesundheits surveys „Health Behaviour in School-aged Children“* (S. 168–189). Weinheim: Beltz Juventa.
- Blömeke, S. (2004). Empirische Befunde zur Wirksamkeit der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 59–91). Bad Heilbrunn & Braunschweig: Klinkhardt & Westermann.
- Blömeke, S. (2009). Lehrerbildung in Deutschland. *PÄD Forum: unterrichten erziehen* 37 (28), 5–8.
- Boekaerts, M. (1998). Do culturally rooted self-construals affect students' conceptualization of control over learning? *Educational Psychologist*, 33, 87–108.
- Bohl, T., Batzel, A. & Richey, P. (2012). Öffnung – Differenzierung – Individualisierung – Adaptivität. Charakteristika, didaktische Implikationen und Forschungsbefunde verwandter Unterrichtskonzepte zum Umgang mit Heterogenität. In T. Bohl, M. Bönsch, M. Trautmann & B. Wischer (Hrsg.), *Binnendifferenzierung. Teil 1: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 40–72). Immenhausen bei Kassel: Prolog.
- Bohl, T., Bönsch, M., Trautmann, M. & Wischer, B. (2012). Binnendifferenzierung – Ein altes Thema in der aktuellen Diskussion. Zur Einleitung. In T. Bohl, M. Bönsch, M. Trautmann & B. Wischer (Hrsg.), *Binnendifferenzierung. Teil 1: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 5–7). Immenhausen bei Kassel: Prolog.
- Bong, M. (2004). Academic Motivation in Self-Efficacy, Task Value, Achievement Goal Orientations, and Attributional Beliefs. *The Journal of Educational Research*, 97 (6), 287–298.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. überarbeitete Aufl.). Berlin: Springer.
- Bös, N. (2013, 5. Februar). Bloß nicht verzetteln. *Frankfurter Allgemeine Zeitung online*. Zugriff am 08.04.2013. Verfügbar unter <http://www.faz.net/aktuell/beruf-chance/campus/lerntechniken-bloss-nicht-verzetteln-12046737.html>

- Bos, W. & Wendt, H. (2013). *TIMSS 2015*. Institut für Schulentwicklungsforschung, TU Dortmund. Zugriff am 16.05.2014. Verfügbar unter <http://ifs-dortmund.de/8602.html>
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.). (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bradley, R. T., McCraty, R., Atkinson, M., Tomasino, D., Daugherty, A. & Arguelles, L. (2010). Emotion Self-Regulation, Psychophysiological Coherence, and Test Anxiety: Results from an Experiment Using Electrophysiological Measures. *Applied Psychophysiology and Biofeedback*, 35 (4), 261–283.
- Brohm, M. (2011). *Motivation lernen. Das Trainingsprogramm für die Schule; mit Übungen und Kopiervorlagen* (1. Aufl.). Weinheim: Beltz.
- Buchwald, P. (2011). *Stress in der Schule und wie wir ihn bewältigen*. Paderborn: Schöningh.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Burkard, C. & Eikenbusch, G. (2007). *Praxishandbuch Evaluation in der Schule* (6. Aufl.). Berlin: Cornelsen Scriptor.
- Carter, E. W., Wehby, J., Hughes, C., Johnson, S. M., Plank, D. R., Barton-Arwood, S. M. et al. (2005). Preparing Adolescents with High-Incidence Disabilities for High-Stakes Testing with Strategy Instruction. *Preventing School Failure*, 49, 55–62.
- Carver, C. S., Scheier, M. F. & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, 30 (7), 879–889.
- Chott, P. O. (1999). Ansätze zur Förderung einer „Fehlerkultur“. Lernförderung in der Schule durch Fehlerprophylaxe und Fehlermanagement. *PÄDForum*, 3, 238–248.
- Chott, P. O. (2004). Ansätze zur Förderung einer „Fehlerkultur“. Das Lernen fördern durch Fehlerprophylaxe und Fehlermanagement. *Lernchancen*, 39 (7), 53–56.
- Chou, C.-P., Yand, D., Pentz, M. A. & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis*, 46, 213–225.
- Chua, Y. P., Tie, F. H. & Don, Z. M. (2011). Creating an Education Research Acculturation Theory for Research Implementation in School. *Education and Urban Society*, 45 (4), 506–526.
- Cohen, M. P. (1998). Determining Sample Sizes for Surveys with Data Analyzed by Hierarchical Linear Models. *Journal of Official Statistics*, 14 (3), 267–275.
- Dambeck, H. (2013, 20. Januar). Erfolg in Mathe: Motivation ist wichtiger als Intelligenz. *Der Spiegel online*. Zugriff am 08.04.2013. Verfügbar unter <http://www.spiegel.de/wissenschaft/mensch/erfolg-in-mathe-motivation-ist-wichtiger-als-intelligenz-a-878609.html>
- DeAnda, D., Baroni, S., Boskin, L., Buchwald, L., Morgan, J., Ow, J. et al. (2000). Stress, stressors and coping among high school students. *Children and Youth Services Review*, 22 (6), 441–463.
- DeCharms, R. (1979). *Motivation in der Klasse*. München: mVg.

- DeCharms, R. (2011). *Motivation in der Klasse. Standardwerke aus der Psychologie und Pädagogik – Reprints* (Bd. 5). Münster, New York: Waxmann.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39, 223–238.
- Deci, E. L. & Ryan, R. M. (2009). Promoting Self-Determined School Engagment. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 172–195). New York: Routledge.
- Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R. et al. (2009). Multilevel Modeling: A Review of Methodological Issues and Applications. *Review of Educational Research*, 79 (1), 69–102.
- DeGEval – Deutsche Gesellschaft für Evaluation e. V. (2008). *Standards für Evaluation*. Mainz: DeGEval.
- DeLongis, A., Coyne, J. C., Dakof, G., Folkman, S. & Lazarus, R. S. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health Psychology*, 1 (2), 119–136.
- Dignath, C., Buettner, G. & Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review*, 3 (2), 101–129.
- Ditton, H. (1998). *Mehrebenenanalyse*. Weinheim: Juventa.
- Domitrovich, C. E. & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, 11, 193–221.
- Drössler, S. (2009). *Förderung sozialer Kompetenzen im Unterricht - Prozessorientierte Evaluation eines Fortbildungsprojekts für die Sekundarstufe I*. Dissertation, Humboldt-Universität zu Berlin.
- Drössler, S., Jerusalem, M. & Mittag, W. (2007). Förderung sozialer Kompetenzen im Unterricht. Implementation eines Lehrerfortbildungsprogramms. *Zeitschrift für pädagogische Psychologie*, 21 (2), 157–168.
- Drössler, S., Röder, B. & Jerusalem, M. (2007). Förderung von Selbstwirksamkeit und Selbstbestimmung im Unterricht. In M. Landmann & B. Schmitz (Hrsg.), *Selbstregulation erfolgreich fördern. Praxisnahe Trainingsprogramme für effektives Lernen* (S. 206–231). Stuttgart: Kohlhammer.
- Durlak, J. A. (2013). *The importance of quality implementation for research, practice, and policy*. ASPE Research Brief – US Department of Health and Human Services.
- Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41 (3-4), 321–350.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D. & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, 82 (1), 405–432.

- Eder, F. (2010). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 694–703). Weinheim: Beltz.
- Egbochuku, E. O. & Obodo, B. O. (2005). Effects of Systematic Desensitisation (SD) Therapy on the Reduction of Test Anxiety Among Adolescents in Nigerian Schools. *Journal of Instructional Psychology*, 32 (4), 298–304.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden. Lehrbuch; mit Online-Materialien* (3. Aufl.). Weinheim: Beltz.
- Eisenträger, U. (2000). Dicht gedrängt. *Erziehung & Wissenschaft*, 12, 32.
- Embse, N. von der, Barterian, J. & Segool, N. (2013). Test Anxiety Interventions for Children and Adolescents: A Systematic Review of Treatment Studies from 2000-2010. *Psychology in the Schools*, 50 (1), 57–71.
- Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multi-level models: A new look at an old issue. *Psychological Methods*, 12 (2), 121–138.
- Endres, W. (2012). *Lerntrainer Motivation. 50 Schritte zu mehr Lust am Lernen; 5. bis 9. Klasse*. Weinheim: Beltz.
- Epstein, J. L. (1988). Effective schools or effective students: Dealing with diversity. In R. Haskins & D. MacRae (Hrsg.), *Policies for America's public schools: Teacher equity indicators* (S. 89-126). Norwood, NJ: Ablex.
- Eschenbeck, H., Lohaus, A. & Kohlmann, C.-W. (2007). Instrumente zur Erfassung von Stress und Coping im Kindesalter. In I. Seiffge-Krenke & A. Lohaus (Hrsg.), *Stress und Stressbewältigung im Kindes- und Jugendalter* (S. 31–46). Göttingen: Hogrefe.
- Escher, F. & Seiffge-Krenke, I. (2013). Coping mit Alltagsstress in verschiedenen Problembereichen. Vergleich klinisch auffälliger und gesunder Jugendlicher. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 41 (5), 295–307.
- Euler, D. & Sloane, P. F. E. (1998). Implementation als Problem der Modellversuchsforschung. *Unterrichtswissenschaft*, 26, 312–326.
- Felner, R., Seitinger, A. M., Brand, S., Burns, A. & Bolton, N. (2007). Creating Small Learning Communities: Lessons From the Project on High-Performing Learning Communities About “What Works” in Creating Productive, Developmentally Enhancing, Learning Contexts. *Educational Psychologist*, 42 (4), 209–221.
- Felson, R. B. & Reed, M. D. (1986). Reference groups and self-appraisals of academic ability and performance. *Social Psychology Quarterly*, 49, 103–109.
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19 (4), 499–505.
- Fixsen, D. L., Naoom, S. F., Blase, K., Friedman, R. & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, Florida: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication).
- Frädlich, S. G. (2006). *Günter, der innere Schweinehund. Ein tierisches Motivationsbuch* (8. Aufl.). Offenbach am Main: GABAL.
- Fraser, B. J. (1982). Individualized Classroom Environment Questionnaire. *American Journal of Evaluation*, 3 (2), 72–73.

- Fraser, B. J. & Walberg, H. J. (1991). *Educational environments. Evaluation, antecedents, and consequences*. Oxford: Pergamon.
- Frenzel, A., Götz, T., Pekrun, R. & Fend, H. (2008). Kontroll-Wert-Modell der Prüfungsangst. In J. M. H. Zumbach (Hrsg.), *Pädagogische Psychologie in Theorie und Praxis* (S. 275–284). Göttingen: Hogrefe.
- Fridrici, M. & Lohaus, A. (2009). Stress-prevention in secondary schools: online- versus face-to-face-training. *Health Education*, 109 (4), 299–313.
- Friedl, S., Huber, L., Kempster, U., Resch, C., Samhaber, E., Schmid, F. et al. (2010). *FAQs zur Begabungs- und Begabtenförderung. Die häufigsten Fragen in Zusammenhang mit (Hoch)Begabung*. Salzburg: ÖZBF – Österreichisches Zentrum für Begabtenförderung und Begabungsforschung.
- Frieling, E. & Sonntag, K. (1999). *Lehrbuch Arbeitspsychologie* (2. Aufl.). Bern: Huber.
- Fries, S. & Souvignier, E. (2009). Training. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 406–428). Heidelberg: Springer.
- Gallagher, M. W. (2012). Self-Efficacy. In V. S. Ramachandran (Hrsg.), *Encyclopedia of human behavior* (2. Aufl., S. 314–320). London: Elsevier/Academic Press.
- Garson, G. D. (2013a). Fundamentals of Hierarchical Linear and Multilevel Modeling. In G. D. Garson (Hrsg.), *Hierarchical linear modeling. Guide and applications* (S. 3–25). Thousand Oaks: Sage Publications.
- Garson, G. D. (Hrsg.). (2013b). *Hierarchical linear modeling. Guide and applications*. Thousand Oaks: Sage Publications.
- Garson, G. D. (2013c). Preparing to Analyze Multilevel Data. In G. D. Garson (Hrsg.), *Hierarchical linear modeling. Guide and applications* (S. 27–53). Thousand Oaks: Sage Publications.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1 (3), 515–533.
- Gollwitzer, M. & Jäger, R. S. (2009). *Evaluation kompakt* (1. Aufl.). Weinheim: Beltz, PVU.
- Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Hrsg.), *Handbook of motivation and cognition: Foundations of social behavior* (S. 53–92). New York: Guilford Press.
- Gräsel, C. & Parchmann, I. (2004). Implementationsforschung – oder: der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft*, 32, 196–214.
- Greenberg, D. F. & Phillips, J. A. (2013). Hierarchical Linear Modeling of Growth Curve Trajectories Using HLM. In G. D. Garson (Hrsg.), *Hierarchical linear modeling. Guide and applications* (S. 249–272). Thousand Oaks: Sage Publications.
- Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P. & Kyriakidou, O. (2004). Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank Quarterly*, 82 (4), 581–629.
- Gregor, A. (2005). Examination Anxiety: Live With It, Control It Or Make It Work For You? *School Psychology International*, 26 (5), 617–635.

- Grimshaw, J., Eccles, M., Thomas, R., MacLennan, G., Ramsay, C., Fraser, C. et al. (2006). Toward Evidence-Based Quality Improvement. *Journal of General Internal Medicine*, 21 (2), 14–20.
- Hager, W. (2000). Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 180–201). Bern: Huber.
- Hager, W. (2008). Evaluation pädagogisch-psychologischer Interventionsmaßnahmen. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 721–732). Göttingen: Hogrefe.
- Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen Sie bewirken können? In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 41–85). Bern: Huber.
- Hampel, P. (2007). Stressbewältigungstrainings im Kindesalter. In I. Seiffge-Krenke & A. Lohaus (Hrsg.), *Stress und Stressbewältigung im Kindes- und Jugendalter* (S. 235–246). Göttingen: Hogrefe.
- Hampel, P., Jahr, A. & Backhaus, O. (2008). Geschlechtsspezifisches Anti-Stress-Training in der Schule. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 57, 20–38.
- Hampel, P., Kümmel, U. & Meier, M. (2009). Schul-basiertes Stressbewältigungstraining in der frühen Adoleszenz. *Zeitschrift für Gesundheitspsychologie*, 17 (1), 1–12.
- Hampel, P. & Petermann, F. (2003). *Anti-Stress-Training für Kinder* (2. Aufl.). Weinheim: Beltz, PVU.
- Han, S. S. & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*, 33 (6), 665–679.
- Hasselhorn, M., Köller, O., Maaz, K. & Zimmer, K. (2014). Implementation wirksamer Handlungskonzepte im Bildungsbereich als Forschungsaufgabe. *Psychologische Rundschau*, 65 (3), 140–149.
- Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analysis relating to achievement*. London: Routledge.
- HBSC Team Deutschland. (2010). *Health Behaviour in School-aged Children (HBSC). Die internationale Kinder- und Jugendgesundheitsstudie in Zusammenarbeit mit der Weltgesundheitsorganisation (WHO)*. Bielefeld: WHO Collaborating Centre for Child and Adolescent Health Promotion. Zugriff am 31.12.2013. Verfügbar unter <http://hbsc-germany.de>
- HBSC Team Deutschland. (2011a). *Studie Health Behaviour in School-aged Children – Faktenblatt „Schulische Belastung von Kindern und Jugendlichen“*. Bielefeld: WHO Collaborating Centre for Child and Adolescent Health Promotion.
- HBSC Team Deutschland. (2011b). *Studie Health Behaviour in School-aged Children – Faktenblatt „Subjektive Beschwerdelast von Kindern und Jugendlichen“*. Bielefeld: WHO Collaborating Centre for Child and Adolescent Health Promotion.
- Heckhausen, H. (1977). Motivation: Kognitionspsychologische Aufspaltung eines summarischen Konstrukts. *Psychologische Rundschau*, 28, 175–189.
- Heckhausen, H. (2010). Vorwort. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (4. Aufl., S. IX–X). Berlin: Springer.

- Heckhausen, H., Gollwitzer, P. M. & Weinert, F. E. (Hrsg.). (1987). *Jenseits des Rubikon. Der Wille in den Humanwissenschaften*. Berlin: Springer.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (5. Aufl.). Seelze-Velber: Kallmeyer in Verbindung mit Klett.
- Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research*, 58, 47–77.
- Hobfoll, S. E. (1988). *The ecology of stress*. New York: Hemisphere Pub. Corp.
- Hobfoll, S. E. (1998). *Stress, culture, and community. The psychology and philosophy of stress*. New York: Plenum Press.
- Hodapp, V., Laux, L. & Spielberger, C. D. (1982). Theorie und Messung der emotionalen und kognitiven Komponente der Prüfungsangst. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 169–184.
- Holmes, T. H. & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, 11 (2), 213–218.
- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd). New York: Routledge.
- Hudley, C., Graham, S. & Taylor, A. (2007). Reducing Aggressive Behavior and Increasing Motivation in School: The Evolution of an Intervention to Strengthen School Adjustment. *Educational Psychologist*, 42 (4), 251–260.
- International Association for the Evaluation of Educational Achievement. (2011). *TIMSS 2011. Trends in International Mathematics and Science Study 2011*. Zugriff am 16.05.2014. Verfügbar unter [http://www.iea.nl/timss\\_2011.html](http://www.iea.nl/timss_2011.html)
- Jacobs, B. (1987). Die Auswirkungen transparenzschaffender Maßnahmen auf die aktuelle Angst von einer Klassenarbeit – Eine Metaanalyse zum Saarbrücker Schulangstprojekt. *Empirische Pädagogik*, 1, 139–160.
- Jacobs, B. (1997). *Angstreduktion in der Prüfung durch transparenzschaffende Maßnahmen?*. Medienzentrum der Philosophischen Fakultät der Universität des Saarlandes. Zugriff am 15.09.2013. Verfügbar unter <http://www.phil.uni-sb.de/~jakobs/trans/index.html>
- Jacobs, B. & Bedersdorfer, H. W. (1984). Transparenzschaffende Maßnahmen zum Abbau von Angst in der Prüfung. *Unterrichtswissenschaft*, 2, 182–197.
- Jacobs, B., Bedersdorfer, H. W. & Bohse-Wagner, N. (1983). *Angstabbau durch Transparenz. Ein Feldexperiment zur ökologischen Validierung einer theoriegeleiteten Intervention zur Reduktion von Angst in der Prüfung*. Saarbrücken: Universität des Saarlandes.
- Jacobson, E. (2011). *Entspannung als Therapie. Progressive Relaxation in Theorie und Praxis* (Bd. 69, 7. Aufl.). Stuttgart: Klett-Cotta.
- Jagacinski, C. M. (1992). The effects of task involvement and ego involvement on achievement-related cognitions and behaviors. In D. H. Schunk & J. L. Meece (Hrsg.), *Student perceptions in the classroom* (S. 307–326). Hillsdale, New Jersey: Lawrence Erlbaum.

- Jekauc, D., Völkle, M., Lämmle, L. & Woll, A. (2012). Fehlende Werte in sportwissenschaftlichen Untersuchungen. *Sportwissenschaft*, 42 (2), 126–136.
- Jerusalem, M. (1990). *Persönliche Ressourcen, Vulnerabilität und Streßerleben*. Göttingen: Hogrefe.
- Jerusalem, M. (2003). Prävention in Schulen. In M. Jerusalem & H. Weber (Hrsg.), *Psychologische Gesundheitsförderung. Diagnostik und Prävention* (461–477). Göttingen: Hogrefe.
- Jerusalem, M. (2005a). *Ergebnisbericht zum Kooperationsprojekt SIGIS – Sicher und gesund in der Schule. Verantwortlich und kompetent Handeln – Schulklima und Schulqualität fördern*. Berlin: Humboldt-Universität zu Berlin, Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie.
- Jerusalem, M. (2005b). Selbstwirksamkeit. In H. Weber (Hrsg.), *Handbuch der Persönlichkeitspsychologie und differentiellen Psychologie* (S. 438–445). Göttingen: Hogrefe.
- Jerusalem, M. (2006). Theoretische Konzeptionen der Gesundheitsförderung im Kindes- und Jugendalter. In A. Lohaus, M. Jerusalem & J. Klein-Heßling (Hrsg.), *Gesundheitsförderung im Kindes- und Jugendalter* (S. 31–57). Göttingen: Hogrefe.
- Jerusalem, M. (2008). *Ergebnisbericht zum Kooperationsprojekt FOSIS – Förderung von Schutzfaktoren in der Schule*. Berlin: Humboldt-Universität zu Berlin, Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie.
- Jerusalem, M., Drössler, S., Kleine, D., Klein-Heßling, J., Mittag, W. & Röder, B. (2007). *Selbstwirksamkeit und Selbstbestimmung im Unterricht*. Berlin: Humboldt-Universität zu Berlin, Lehrstuhl für Pädagogische Psychologie und Gesundheitspsychologie.
- Jerusalem, M. & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung. In R. Schwarzer & M. Jerusalem (Hrsg.), *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* (S. 18–19). Berlin: Freie Universität Berlin.
- Jerusalem, M. & Schwarzer, R. (1991). Entwicklung des Selbstkonzepts in verschiedenen Lernumwelten. In R. Pekrun & H. Fend. (Hrsg.), *Schule und Persönlichkeitsentwicklung. Ein Resümee der Längsschnittforschung* (S. 115–128). Stuttgart: Enke.
- Johnson, D. W. & Johnson, R. T. (1995). Positive interdependence: Key to effective cooperation. In R. Hertz-Lazarowitz & N. Miller (Hrsg.), *Interaction in cooperative groups. The theoretical anatomy of group learning* (S. 174–199). Cambridge, New York: Cambridge University Press.
- Kalafat, J., Illback, R. J. & Sanders, D. (2007). The relationship between implementation fidelity and educational outcomes in a school-based family support program: development of a model for evaluating multidimensional full-service programs. *Evaluation and Program Planning*, 30 (2), 136–148.
- Kaluza, G. (2011). *Stressbewältigung. Trainingsmanual zur psychologischen Gesundheitsförderung* (2. Aufl.). Berlin: Springer.
- Kanner, A. D., Coyne, J. C., Schaefer, C. & Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine*, 4 (1), 1–39.
- Kauß, F. (2013, 11. Januar). Was Mitarbeiter wirklich motiviert. *ZEIT online*. Zugriff am 10.02.2015. Verfügbar unter <http://www.zeit.de/karriere/beruf/2012-12/personalfuehrung-motivation-mitarbeiter>



- Kirkpatrick, D. L. (1998). *Evaluating training programs: the four levels* (2. Aufl.). San Francisco: Berett-Koehler.
- Klassen, R. (2002). Writing in Early Adolescence: A Review of the Role of Self-Efficacy Beliefs. *Educational Psychology Review*, 14 (2), 173–203.
- Kleiber, D. (2012). Prävention und Gesundheitsförderung. In A. Auckenthaler (Hrsg.), *Kurzlehrbuch Klinische Psychologie und Gesundheitspsychologie – Grundlagen, Praxis, Kontext*. S. 120-147. Stuttgart: Thieme.
- Klein-Heßling, J. & Lohaus, A. (2012). *Stresspräventionstraining für Kinder im Grundschulalter* (3. Aufl.). Göttingen: Hogrefe.
- Kolbe, M., Jerusalem, M. & Mittag, W. (1998). Veränderungen von Selbstwirksamkeit und Klassenklima im zeitlichen Verlauf. *Unterrichtswissenschaft*, 26 (2), 116–126.
- Köller, O. (2009). Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 333–352). Heidelberg: Springer.
- König, J. (2009). Klassenklima und schulbezogene Hilfslosigkeit in den Jahrgangsstufen 8 und 9. *Zeitschrift für pädagogische Psychologie*, 23 (1), 41–52.
- Kraag, G., Zeegers, M. P., Kok, G., Hosman, C. & Abu-Saad, H. H. (2006). School programs targeting stress management in children and adolescents: A meta-analysis. *Journal of School Psychology*, 44 (6), 449–472.
- Krapp, A. & Ryan, R. (2002). Selbstwirksamkeit und Lernmotivation. Eine kritische Betrachtung der Theorie von Bandura aus der Sicht der Selbstbestimmungstheorie und der pädagogisch-psychologischen Interessentheorie. In M. Jerusalem & D. Hopf (Hrsg.), *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen*, *Zeitschrift für Pädagogik*, 44. Beiheft (S. 54–82). Weinheim: Beltz.
- Kress, K. & Pappas, M. (2013). *Binnendifferenzierung in der Sekundarstufe – das Praxisbuch. Profi-Tipps und Materialien aus der Lehrerfortbildung [Sekundarstufe I + II; Methodenkartei, Vorlagen, Checklisten]* (1. Aufl.). Donauwörth: Auer.
- Kröniger-Jungaberle, H. & Grevenstein, D. (2013). Development of salutogenetic factors in mental health – Antonovsky’s sense of coherence and Bandura’s self-efficacy related to Derogatis’ symptom check list (SCL-90-R). *Health and Quality of Life Outcomes*, 11 (1), 80.
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts* (1. Aufl.). Paderborn: UTB.
- Landmann, M., Perels, F., Otto, B. & Schmitz, B. (2009). Selbstregulation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 49–70). Heidelberg: Springer.
- Lang, J. W. & Lang, J. (2010). Priming Competence Diminishes the Link Between Cognitive Test Anxiety and Test Performance: Implications for the Interpretation of Test Scores. *Psychological Science*, 21 (6), 811–819.
- Larson, H. A., El Ramahi, M. K., Conn, S. R., Estes, L. A. & Ghibellini, A. B. (2010). Reducing Test Anxiety Among Third Grade Students Through the Implementation of Relaxation Techniques. *Journal of School Counseling*, 8 (19), 1–19.
- Laux, L. (1983). Psychologische Streßkonzeptionen. In H. Thoma (Hrsg.), *Theorien und Formen der Motivation* (S. 453–535). Göttingen: Hogrefe.

- Lavasani, M. G., Mirhosseini, F. S., Hejazi, E. & Davoodi, M. (2011). The Effect of Self-regulation Learning Strategies Training on the Academic Motivation and Self-efficacy. *Procedia - Social and Behavioral Sciences*, 29, 627–632.
- Lazarus, R. S. (1999). *Stress and emotion. A new synthesis*. London: Free Association Books.
- Lazarus, R. S. & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Lazarus, R. S. & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of Personality*, 1 (3), 141–169.
- Lazarus, R. S. & Launier, R. (1978). Stress-related transaction between person and environment. In L. A. Pervin & M. Lewis (Hrsg.), *Perspectives in Interactional Psychology* (S. 287–327). Boston, MA: Springer US.
- Lee, W., Lee, M.-J. & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology*, 39 (2), 86–99.
- Lewin, K. (1951). *Field theory in social science*. Chicago: University of Chicago Press.
- Liebert, R. M. & Morris, L. L. (1967). Cognitive and emotional components of test anxiety. A distinction and some initial data. *Psychological Reports*, 20, 975–978.
- Linnenbrink, E. A. & Pintrich, P. R. (2002). Motivation as an Enabler for Academic Success. *School Psychology Review*, 31 (3), 313–327.
- Linnenbrink, E. A. & Pintrich, P. R. (2003). The Role of Self-Efficacy Beliefs in Student Engagement and Learning in the Classroom. *Reading and Writing Quarterly*, 19, 119–137.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83 (404), 1198–1202.
- Lohaus, A. (2006). *Gesundheit und Krankheit aus der Sicht von Kindern* (2. Aufl.). Göttingen: Hogrefe.
- Lohaus, A. (2009). Stressbewältigungskompetenzen. In A. Lohaus (Hrsg.), *Psychologische Förder- und Interventionsprogramme für das Kindes- und Jugendalter* (S. 131–140). Heidelberg: Springer.
- Lohaus, A., Beyer, A. & Klein-Heßling, J. (2004). Stresserleben und Stresssymptomatik bei Kindern und Jugendlichen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36 (1), 38–46.
- Lohaus, A., Domsch, H. & Fridrici, M. (2007). *Stressbewältigung für Kinder und Jugendliche. [positiv mit Stress umgehen lernen; konkrete Tipps und Übungen; Hilfen für Eltern und Lehrer]*. Heidelberg: Springer.
- Lohaus, A., Eschenbeck, H., Kohlmann, C.-W. & Klein-Heßling, J. (2006). *Fragebogen zur Erhebung von Stress und Stressbewältigung im Kindes- und Jugendalter (SSKJ 3-8)*. Göttingen: Hogrefe.
- Lohaus, A., Fridrici, M. & Maass, A. (2009). Stressprävention im Jugendalter. *Zeitschrift für Gesundheitspsychologie*, 17 (1), 13–21.

- Lou, Y., Abrami, P. C. & Spence, J. C. (2000). Effects of Within-Class Grouping on Student Achievement: An Exploratory Model. *The Journal of Educational Research*, 94 (2), 101–112.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58 (2), 103–117.
- Luhmann, M. & Eid, M. (2013). Studying Reaction to Repeated Life Events With Discontinuous Change Models Using HLM. In G. D. Garson (Hrsg.), *Hierarchical linear modeling. Guide and applications* (S. 273–289). Thousand Oaks: Sage Publications.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks: Sage Publications.
- Lyrakos, D. G. (2012). The Impact of Stress, Social Support, Self-Efficacy and Coping on University Students, a Multicultural European Study. *Psychology*, 3 (2), 143–149.
- Maddux, J. E. (1995). Self-Efficacy-Theory: An Introduction. In J. E. Maddux (Hrsg.), *Self-efficacy, adaption, and adjustment: Theory, research, and application* (S. 3–33). New York: Plenum Press.
- Maerten-Rivera, J. (2013). A Piecewise Growth Model Using HLM 7 to Examine Change in Teaching Practices Following a Science Teacher Development Intervention. In G. D. Garson (Hrsg.), *Hierarchical linear modeling. Guide and applications* (S. 249–272). Thousand Oaks: Sage Publications.
- Magadi, M. & Desta, M. (2011). A multilevel analysis of the determinants and cross-national variations of HIV seropositivity in sub-Saharan Africa: Evidence from the DHS. *Health & Place*, 17 (5), 1067–1083.
- Mandler, G. & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 47, 166–173.
- Margolis, H. & McCabe, P. P. (2006). Improving Self-Efficacy and Motivation: What to do, What to say. *Intervention in School and Clinic*, 41 (4), 218–227.
- Marsh, H. W. (2005). Big-Fish-Little-Pond Effect on Academic Self-Concept. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology*, 19 (3), 119–127.
- Martin, A. J. (2008). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology*, 33 (2), 239–269.
- McClelland, D. C., Atkinson, J. W., Clark, R. A. & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts.
- Meichenbaum, D. (2012). *Intervention bei Stress. Anwendung und Wirkung des Stressimpfungstrainings* (3. Aufl.). Bern: Huber.
- Meyers, D. C., Durlak, J. A. & Wandersman, A. (2012). The Quality Implementation Framework: A Synthesis of Critical Steps in the Implementation Process. *American Journal of Community Psychology*, 50 (3-4), 462–480.
- Michie, S., Fixsen, D., Grimshaw, J. M. & Eccles, M. P. (2005). Making psychological theory useful for implementing evidence based practice: a consensus approach. *Quality and Safety in Health Care*, 14 (1), 26–33.

- Michie, S., Fixsen, D., Grimshaw, J. M. & Eccles, M. P. (2009). Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science*, 4 (1), 40.
- Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 102–128). Bern: Huber.
- Moeini, B., Froug, S., Hidarnia, A., Babaii, G. R., Birashk, B. & Allahverdipour, H. (2008). Perceived stress, self-efficacy and its relations to psychological well-being status in Iranian male high school students. *Social Behavior and Personality*, 36 (2), 257–266.
- Möller, J. (2010). Attributionen. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 38–45). Weinheim: Beltz.
- Mühlig, S. & Poldrack, A. (2011). Kognitive Therapieverfahren. In H.-U. Wittchen (Hrsg.), *Klinische Psychologie & Psychotherapie* (2. Aufl., S. 543–564). Heidelberg: Springer-Medizin.
- Muthén, B. (2001). Second-generation structural equation modeling with varying parameters with a combination of categorical and continuous latent variables: New opportunities for latent class latent growth modeling. In L. M. Collins & A. Sayer (Hrsg.), *New methods for the analysis of change* (1. Aufl., S. 291–322). Washington, DC: American Psychological Association.
- Nater, U. M., Ditzen, B. & Ehler, U. (2011). Stressabhängige körperliche Beschwerden. In H.-U. Wittchen (Hrsg.), *Klinische Psychologie & Psychotherapie* (2. Aufl., S. 1039–1052). Heidelberg: Springer-Medizin.
- Neber, H. & Fischer, F. (2010). Kooperatives Lernen. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 394–402). Weinheim: Beltz.
- Nezlek, J., Schröder-Abé, M. & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung. *Psychologische Rundschau*, 57, 213–223.
- OECD. (2011). *PISA 2009 Ergebnisse – Lernfortschritte im globalen Wettbewerb. Veränderungen bei den Schülerleistungen seit 2000* (Band 5). OECD Publishing.
- OECD. (2014a). *PISA – Internationale Schulleistungsstudie der OECD*. Zugriff am 16.05.2014. Verfügbar unter <http://www.oecd.org/berlin/themen/pisa-internationaleschulleistungsstudiederoecd.htm>
- OECD. (2014b). *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*. OECD Publishing.
- Oser, F. & Spychiger, M. (2005). *Lernen ist schmerzhaft*. Weinheim: Beltz.
- Øvretveit, J. (2002). *Evaluation gesundheitsbezogener Interventionen. Einführung in die Bewertung von gesundheitsbezogenen Behandlungen, Dienstleistungen, Richtlinien und organisationsbezogenen Interventionen* (1. Aufl.). Bern: Huber.
- Owens, J. S., Lyon, A. R., Brandt, N. E., Warner, C. M., Nadeem, E., Spiel, C. et al. (2014). Implementation Science in School Mental Health: Key Constructs in a Developing Research Agenda. *School Mental Health*, 6 (2), 99–111.
- Pajares, F. (1996). Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research*, 66 (4), 543–578.

- Paradies, L. & Linser, H. J. (2009). Organisationsformen, Differenzierung, Integration. In K.-H. Arnold (Hrsg.), *Handbuch Unterricht* (2. Aufl., S. 261–265). Bad Heilbrunn: Klinkhardt.
- Patry, J.-L. & Hager, W. (2000). Abschließende Bemerkungen: Dilemmata in der Evaluation. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 258–275). Bern: Huber.
- Payne, A. A. (2009). Do predictors of the implementation quality of school-based prevention programs differ by program type? *Prevention Science*, 10 (2), 151–167.
- Payne, A. A. & Eckert, R. (2009). The Relative Importance of Provider, Program, School, and Community Predictors of the Implementation Quality of School-Based Prevention Programs. *Prevention Science*, 11 (2), 126–141.
- Payne, A. A., Gottfredson, D. C. & Gottfredson, G. D. (2006). School Predictors of the Intensity of Implementation of School-Based Prevention Programs: Results from a National Study. *Prevention Science*, 7 (2), 225–237.
- Pekrun, R. (1992). Expectancy-value theory of anxiety: Overview and implications. In D. G. Forgays, Sosnowski T. & Wrzesniewski K. (Hrsg.), *Anxiety: Recent developments in self-appraisal, psychophysiological and health research* (S. 23–41). Washington: Hemisphere.
- Pekrun, R. (2009). Emotions at School. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 575–604). New York: Routledge.
- Pekrun, R. & Götz, T. (2006). Emotionsregulation: Vom Umgang mit Prüfungsangst. In H. Mandl & H. F. Friedrich (Hrsg.), *Handbuch Lernstrategien* (S. 248–258). Göttingen: Hogrefe.
- Perrez, M. & Reicherts, M. (1992). *Stress, coping, and health. A situation behavior approach, - theory, methods, applications*. Seattle: Hogrefe & Huber.
- Petermann, F. (2014a). Implementationsforschung. *Psychologische Rundschau*, 65 (3), 121.
- Petermann, F. (2014b). Implementationsforschung: Grundbegriffe und Konzepte. *Psychologische Rundschau*, 65 (3), 122–128.
- Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology*, 95 (4), 667–686.
- Projektgruppe „Feedback-Materialien“ (Haider, G., Hrsg.). (2013). *Qualität in Schulen – Fragebögen*. Das Zukunftministerium bm:bwk. Zugriff am 10.02.2015. Verfügbar unter <http://www.qis.at/FBGesamt.pdf>
- QuestBack GmbH. (2014). *unipark.info*. Zugriff am 05.01.2014. Verfügbar unter [www.unipark.de](http://www.unipark.de)
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y., Congdon, R. & Du Toit, M. (2004). *HLM 6: Hierarchical linear & nonlinear modeling [Benutzerhandbuch]*. Lincolnwood, IL: Scientific Software International, Inc.

- Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2009) HLM 6.08 für Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.
- Reimann, S. & Hammelstein, P. (2006). Ressourcenorientierte Ansätze. In B. Renneberg & P. Hammelstein (Hrsg.), *Gesundheitspsychologie* (S. 13–28). Berlin: Springer.
- Renkl, A. & Mandl, H. (1995). Kooperatives Lernen: Die Frage nach dem Notwendigen und dem Ersetzbaren. *Unterrichtswissenschaft*, 23, 292–300.
- Resnicow, K., Cohn, L., Reinhardt, J., Cross, D., Futterman, R., Kirschner, E. et al. (1992). A three-year evaluation of the know your body program in inner-city schoolchildren. *Health Education Quarterly*, 19, 463–480.
- Rheinberg, F. (1986). Lernmotivation. In W. F. R. Sarges (Hrsg.), *Psychologie für die Erwachsenenbildung* (S. 360–365). Göttingen: Hogrefe.
- Rheinberg, F. (1989). *Zweck und Tätigkeit*. Göttingen: Hogrefe.
- Rheinberg, F. (1995). Individuelle Bezugsnormen der Leistungsbewertung und Motivation im Unterricht. *Pädagogische Welt*, 49, 59–62.
- Rheinberg, F. (2005). Verhaltenslenkung in Lehr- und Erziehungssituationen. In F. Rheinberg & S. Krug (Hrsg.), *Motivationsförderung im Schulalltag* (S. 15–23). Göttingen: Hogrefe.
- Rheinberg, F. (2010). Intrinsische Motivation und Flow-Erleben. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (4. Aufl., S. 365–387). Berlin: Springer.
- Rheinberg, F. & Krug, S. (Hrsg.). (2005a). *Motivationsförderung im Schulalltag*. Göttingen: Hogrefe.
- Rheinberg, F. & Krug, S. (2005b). Vorwort zur ersten Auflage. In F. Rheinberg & S. Krug (Hrsg.), *Motivationsförderung im Schulalltag* (S. 9–10). Göttingen: Hogrefe.
- Rheinberg, F. & Vollmeyer, R. (2012). *Motivation* (8. Aufl.). Stuttgart: Kohlhammer.
- Robert Koch Institut (RKI). (2015). *KIGGS – Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland*. RKI. Zugriff am 10.02.2015. Verfügbar unter <http://www.kiggs-studie.de>
- Röder, B. (2008). *Selbstwirksamkeitsförderung durch Motivierung von Schülern*. Dissertation, Freie Universität Berlin.
- Röder, B., Drössler, S. & Jerusalem, M. (2010). Selbstwirksamkeit und Selbstbestimmung im Unterricht. In T. Hascher (Hrsg.), *Pädagogische Interventionsforschung. Theoretische Grundlagen und empirisches Handlungswissen* (S. 157–187). Weinheim: Juventa.
- Röder, B. & Jerusalem, M. (2007). Implementationsgrad und Wirkungen eines Programms zur Förderung von Selbstwirksamkeit. *Psychologie in Erziehung und Unterricht*, 54, 30–46.
- Rohmert, W. & Rutenfranz, J. (1975). *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen*. Bonn: Bundesministerium für Arbeit und Sozialordnung, Referat Öffentlichkeitsarbeit.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation. A systematic approach* (7. Aufl.). Thousand Oaks, CA: Sage.
- Rost, D. H. (Hrsg.). (2010). *Handwörterbuch pädagogische Psychologie* (4. Aufl.). Weinheim: Beltz.

- Rost, D. H. (Hrsg.). (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung* (3. Aufl.). Bad Heilbrunn: UTB.
- Rost, D. H. & Schermer, F. J. (2010). Leistungsängstlichkeit. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 451–465). Weinheim: Beltz.
- Rothermund, K. & Brandtstädter, J. (1997). Entwicklung und Bewältigung: Festhalten und Preisgeben von Zielen als Formen der Bewältigung von Entwicklungsproblemen. In C. Tesch-Römer, C. Salewski & G. Schwarz (Hrsg.), *Psychologie der Bewältigung* (S. 121–133). Weinheim: Beltz.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monograph: General and Applied*, 80 (1), 1–28.
- Ryan, R. (2012). Motivation and the Organization of Human Behavior: Three Reasons for the Reemergence of a field. In R. M. Ryan (Hrsg.), *The Oxford handbook of human motivation* (Oxford library of psychology, S. 3–10). New York: Oxford University Press.
- Ryan, R. M. & Legate, N. (2012). Through a Fly's Eye: Multiple Yet Overlapping Perspectives on Future Directions for Human Motivation Research. In R. M. Ryan (Hrsg.), *The Oxford handbook of human motivation* (Oxford library of psychology, S. 554–564). New York: Oxford University Press.
- Saldern, M. & Littig, K. E. (1987). *Landauer Skalen zum Sozialklima (LASSO 4-13)*. Weinheim: Beltz.
- Sarason, S. B., Davidson, K., Lighthall, F. & Waite, R. (1958). A Test Anxiety Scale for Children. *Child Development*, 29 (1), 105–113.
- Satow, L. (1999). *Klassenklima und Selbstwirksamkeitsentwicklung*. Dissertation, Freie Universität Berlin.
- Satow, L. (2002). Unterrichtsklima und Selbstwirksamkeitsdynamik. In M. Jerusalem & D. Hopf (Hrsg.), *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen, Zeitschrift für Pädagogik*, 44. Beiheft (S. 174–191). Weinheim: Beltz.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147–177.
- Scheffer, D. & Heckhausen, H. (2010). Eigenschaftstheorien der Motivation. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (4. Aufl., S. 43–72). Berlin: Springer.
- Scheier, M. F. & Carver, C. S. (1987). Dispositional Optimism and Physical Well-Being: The Influence of Generalized Outcome Expectancies on Health. *Journal of Personality and Social Psychology*, 55 (2), 169–210.
- Scheirer, M. A., Shediak, M. C. & Cassady, C. E. (1995). Measuring the implementation of health promotion programs: the case of the Breast and Cervical Cancer Program in Maryland. *Health Education Research*, 10, 11–25.
- Schiefele, U. (2009). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 152–177). Heidelberg: Springer.
- Schmidt, D. (2010). *Motivation. 88 Strategien, Impulse und Tipps für eine hohe Selbstmotivation*: Gabler.
- Schnack, J. (1997). Evaluation – Eine Basisbibliothek. Die wichtigsten deutschsprachigen Veröffentlichungen zum Thema. *Pädagogik*, 49 (5), 35–37.

- Schneider, C. & Ludwig, P. H. (2012). Auswirkungen von Maßnahmen der inneren Leistungsdifferenzierung auf Schulleistung und Fähigkeitsselbstkonzept im Vergleich zur äußeren Differenzierung. In T. Bohl, M. Bönsch, M. Trautmann & B. Wischer (Hrsg.), *Binnendifferenzierung. Teil 1: Didaktische Grundlagen und Forschungsergebnisse zur Binnendifferenzierung im Unterricht* (S. 72–106). Immenhausen bei Kassel: Prolog.
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26, 207–231.
- Schunk, D. H. (1995a). Self-efficacy and education and instruction. In J. E. Maddux (Hrsg.), *Self-efficacy, adaption, and adjustment: Theory, research, and application* (S. 281–303). New York: Plenum Press.
- Schunk, D. H. (1995b). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, 7 (2), 112–137.
- Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading and Writing Quarterly*, 19, 159–172.
- Schunk, D. H. & Pajares, F. (2009). Self-Efficacy Theory. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 35–53). New York: Routledge.
- Schwarzer, R. (2000). *Stress, Angst und Handlungsregulation* (4. Auflage). Stuttgart: Kohlhammer.
- Schwarzer, R. (2004). *Psychologie des Gesundheitsverhaltens. Einführung in die Gesundheitspsychologie* (3. Aufl.). Göttingen: Hogrefe.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. In M. Jerusalem & D. Hopf (Hrsg.), *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen, Zeitschrift für Pädagogik*, 44. Beiheft (S. 28–53). Weinheim: Beltz.
- Seiffge-Krenke, I. (2007). Bewältigung von familiären und schulischen Problemen. In I. Seiffge-Krenke & A. Lohaus (Hrsg.), *Stress und Stressbewältigung im Kindes- und Jugendalter* (S. 161–174). Göttingen: Hogrefe.
- Seiffge-Krenke, I. (2008). Schulstress in Deutschland: Ursachen, Häufigkeit und internationale Verortung. *Praxis der Kinderpsychologie und Kinderpsychiatrie* (57), 3–19.
- Seiffge-Krenke, I., Gelhaar, T. & Kollmar, F. (2007). Instrumente zur Erfassung von Stress und Coping im Jugendalter. In I. Seiffge-Krenke & A. Lohaus (Hrsg.), *Stress und Stressbewältigung im Kindes- und Jugendalter* (S. 47–65). Göttingen: Hogrefe.
- Seiffge-Krenke, I., Persike, M., Chau, C., Hendry, L. B., Kloeppe, M., Terzini-Hollar, M. et al. (2012). Differences in agency? How adolescents from 18 countries perceive and cope with their futures. *International Journal of Behavioral Development*, 36 (4), 258–270.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development and death*. San Francisco: Freeman.
- Selye, H. (1936). A Syndrome produced by Diverse Nocuous Agents. *Nature*, 138 (3479), 32.
- Selye, H. (1978). *The Stress of Life*. New York: McGraw-Hill.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft Berlin. (2014). *Schulreform. Ergebnisse und Ausblick*. Zugriff am 06.01.2014. Verfügbar unter <http://www.berlin.de/sen/bildung/bildungspolitik/schulreform/>



- Senatsverwaltung für Integration, Arbeit und Soziales Berlin. (2011). *Berliner Masterplan Qualifizierung. Ganzheitlich – Nachhaltig – Miteinander*, Berlin. Zugriff am 05.01.2015. Verfügbar unter [http://www.berlin.de/imperia/md/content/sen-arbeit/besch-impulse/lernen/berliner\\_masterplan\\_qualifizierung.pdf?start&ts=1405412891&file=berliner\\_masterplan\\_qualifizierung.pdf](http://www.berlin.de/imperia/md/content/sen-arbeit/besch-impulse/lernen/berliner_masterplan_qualifizierung.pdf?start&ts=1405412891&file=berliner_masterplan_qualifizierung.pdf)
- Shen, Y. E. (2009). Relationships between self-efficacy, social support and stress coping strategies in Chinese primary and secondary school teachers. *Stress and Health*, 25 (2), 129–138.
- Siegle, D. & McCoach, D. B. (2007). Increasing Student Mathematics Self-Efficacy Through Teacher Training. *Journal of Advanced Academics*, 18 (2), 278–312.
- Silbereisen, R. K. & Weichold, K. (2012). Jugend (12-19 Jahre). In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie. Mit Online-Materialien* (7. Aufl., S. 235–258). Weinheim: Beltz.
- Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: University Press.
- Slavin, R. E. (1989). Cooperative Learning and student achievement. In R. E. Slavin (Hrsg.), *School and classroom organization* (S. 129–156). Hillsdale: L. Erlbaum Associates.
- Slavin, R. E. (1995). *Cooperative learning. Theory, research, and practice* (2. Aufl.). Boston: Allyn and Bacon.
- Snijders, T. A. B. & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18, 237–259.
- Snijders, T. A. B. & Bosker, R. J. (1994). Modeled Variance in two-level models. *Sociological Methods & Research*, 22 (3), 342–363.
- Snyder, J., Bolin, F. & Zuwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Hrsg.), *Handbook of research on curriculum* (S. 402–435). New York: Macmillan.
- Souvignier, E. & Mokhlesgerami, J. (2005). Implementation eines Programms zur Vermittlung von Lesestrategien im Deutschunterricht. *Zeitschrift für pädagogische Psychologie*, 19 (4), 249–261.
- Spielberger, C. D. (1980). *Test Anxiety Inventory: Preliminary Professional Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Spinath, B. (2008). Bezugsnormorientierung. In J. M. H. Zumbach (Hrsg.), *Pädagogische Psychologie in Theorie und Praxis* (S. 185–192). Göttingen: Hogrefe.
- Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2014). *TIMSS*. Zugriff am 16.05.2014. Verfügbar unter <http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsmonitoring/internationale-schulleistungsvergleiche/timss.html>
- Stith, S., Pruitt, I., Dees, J., Fronce, M., Grenn, N., Som, A. et al. (2006). Implementing Community-Based Prevention Programming: A Review of the Literature. *The Journal of Primary Prevention*, 27 (6), 599–617.
- Stockmann, R. (2006). Evaluation in Deutschland. In R. Stockmann (Hrsg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (Bd. 1, 3. Aufl., S. 15–46). Münster, New York: Waxmann.

- Stockmann, R. & Meyer, W. (2014). *Evaluation. Eine Einführung* (2. Aufl.). Opladen: Budrich.
- Streblow, L., Schiefele, U. & Riedel, S. (2012). Überprüfung des revidierten Trainings zur Förderung der Lesekompetenz und der Lesemotivation (LeKoLemo) für die Sekundarstufe I. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44 (1), 12–26.
- Strittmatter, P. (1997). *Schulangstreduktion. Abbau von Angst in schulischen Leistungssituationen*. Luchterhand: Neuwied.
- Suldo, S. M., Shaunessy, E., Thalji, A., Michalowski, J. & Shaffer, E. (2009). Sources of Stress for Students in High School College Preparatory and General Education Programs. Group Differences and Associations With Adjustment. *Adolescence*, 44 (176), 925–948.
- Sun, R. C. F., Shek, D. T. L. & Siu, A. M. H. (2008). Positive school and classroom environment: precursors of successful implementation of positive youth development programs. *ScientificWorldJournal*, 8, 1063–1074.
- Taché, J. & Selye, H. (1985). On stress and coping mechanisms. *Issues in mental health nursing*, 7 (1-4), 3–24.
- Thaker, R. B. & Verma, A. P. (2014). A study of perceived stress and coping styles among mid adolescents. *National Journal of Physiology, Pharmacy and Pharmacology*, 4 (1), 1.
- Traub, S. (2003). Selbstgesteuertes Lernen in der Praxis. *Pädagogik*, 55 (5), 19–22.
- Trautwein, U. & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind – Referenzgruppeneffekte bei Übertrittsentscheidungen. *Zeitschrift für pädagogische Psychologie*, 27 (2), 119–133.
- Ulich, E. (2011). *Arbeitspsychologie* (7. Aufl.). Zürich: Schäffer-Poeschel.
- Usher, E. & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of Literature and Future Directions. *Review of Educational Research*, 78 (4), 751–796.
- Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.
- Wagner, M. (2007). Schulschwänzen und Schulverweigerung: Ergebnisse und Defizite der Forschung. In M. Wagner (Hrsg.), *Schulabsentismus. Soziologische Analysen zum Einfluss von Familie, Schule und Freundeskreis* (S. 239–252). Weinheim: Juventa.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillmann, L. et al. (2008). Bridging the Gap between Prevention Research and Practice: The Interactive Systems Framework for Dissemination and Implementation. *American Journal of Community Psychology*, 41, 171–181.
- Wang, C., Kim, D.-H., Bong, M. & Ahn, H. S. (2013). Examining measurement properties of an English Self-Efficacy scale for English language learners in Korea. *International Journal of Educational Research*, 59, 24–34.
- Wegner, K. (2014, 14. Januar). *Schüler in der Leistungsfalle*. ZDF.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92 (4), 548–573.
- Weiner, B. (1986). *Attributional theory of motivation and emotion*. Heidelberg: Springer.

- Weiner, B., Frieze, I. H., Kulka, A., Reed, L., Rest, S. & Rosenbaum, R. M. (1971). *Perceiving the causes of success and failure*. New York: General Learning Press.
- Weinert, F. (1982). Selbstgesteuertes Lernen als Voraussetzung, Methode und Ziel des Unterrichts. *Unterrichtswissenschaft*, 10, 99–110.
- Weissberg, R. P., Kumpfer, K. & Seligman, M. E. P. (2003). Prevention that works for children and youth: An introduction. *American Psychologist*, 58 (6-7), 425–432.
- Wentzel, K. R. & Wigfield, A. (2009). Introduction. In K. R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 1–8). New York: Routledge.
- Werkstatt. (2001). Gut geregelt. *Lernende Schule*, 14, 33–57.
- Westermann, R. (2002). Merkmale und Varianten von Evaluationen. Überblick und Klassifikation. *Zeitschrift für Psychologie*, 210 (1), 4–26.
- Wigfield, A., Cambria, J. & Eccles, J. S. (2012). Motivation in Education. In R. M. Ryan (Hrsg.), *The Oxford handbook of human motivation* (Oxford library of psychology, S. 463–478). New York: Oxford University Press.
- Wigfield, A., Ho, A. N. & Mason-Singh, A. (2011). Achievement Motivation. In B. B. Brown & M. J. Prinstein (Hrsg.), *Encyclopedia of adolescence* (S. 10–19). Amsterdam: Elsevier/Academic Press.
- Wigfield, A. & Wentzel, K. R. (2007). Introduction to Motivation at School: Interventions That Work. *Educational Psychologist*, 42 (4), 191–196.
- Wild, E., Hofer, M. & Pekrun, R. (2006). Psychologie des Lernens. In A. Krapp & Weidenmann B. (Hrsg.), *Pädagogische Psychologie. Ein Lehrbuch* (5. Aufl., S. 203–268). Weinheim: Beltz.
- Wine, J. D. (1971). Test anxiety and the direction of attention. *Psychological Bulletin*, 76, 92–104.
- Wirtz, M. (Hrsg.). (2014). *Dorsch - Lexikon der Psychologie* (17. überarbeitete Aufl.). Bern: Huber.
- Wottawa, H. & Thierau, H. (2003). *Lehrbuch Evaluation* (3., korrigierte Aufl.). Bern: Huber.
- Wright, S. L., Jenkins-Guarnieri, M. A. & Murdock, J. L. (2013). Career Development Among First-Year College Students: College Self-Efficacy, Student Persistence, and Academic Success. *Journal of Career Development*, 40 (4), 292–310.
- Yankey, T. & Biswas, U. N. (2012). Life Skills Training as an Effective Intervention Strategy to Reduce Stress among Tibetan Refugee Adolescents. *Journal of Refugee Studies*, 25 (4), 514–536.
- Zeidner, M. (1998). *Test anxiety. The state of the art*. New York: Plenum Press.
- Zeidner, M. (2007). Test Anxiety in Educational Contexts: Concepts, Findings, and Future Directions. In P. A. Schutz & R. Pekrun (Hrsg.), *Emotion in education* (S. 165–184). Amsterdam: Academic Press.
- Zhang, Q. (2014). Assessing the Effects of Instructor Enthusiasm on Classroom Engagement, Learning Goal Orientation, and Academic Self-Efficacy. *Communication Teacher*, 28 (1), 44–56.

- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 13–39). San Diego: Academic Press.
- Zumhasch, C. (2009). Das Unterrichtsklima. In K.-H. Arnold (Hrsg.), *Handbuch Unterricht* (2. Aufl., S. 108–111). Bad Heilbronn: Klinkhardt.

## Tabellenverzeichnis

<b>Tabelle 1:</b> Evaluationsbefragungen – Gesamtstichprobe der Schüler .....	126
<b>Tabelle 2:</b> Evaluationsbefragungen – Schülerstichprobe, getrennt nach Interventions- und Kontrollgruppe .....	127
<b>Tabelle 3:</b> Modulspezifische Rückmeldungen – Schülerstichprobe (nur Interventionsschüler). ....	128
<b>Tabelle 4:</b> Analyse-Stichprobe für die Evaluation der Wirksamkeit (T1-T2-T3) .....	137
<b>Tabelle 5:</b> Analyse-Stichprobe für die Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation der Interventionsschüler (T1, MSR1-3, T2) .....	138
<b>Tabelle 6:</b> Lineare Veränderung in der schulischen Selbstwirksamkeitserwartung über beide Projektjahre .....	168
<b>Tabelle 7:</b> Lineare Veränderung in der Prüfungsängstlichkeit über beide Projektjahre...	171
<b>Tabelle 8:</b> Lineare Veränderung in der wahrgenommenen Herausforderung über beide Projektjahre.....	173
<b>Tabelle 9:</b> Lineare Veränderung in der wahrgenommenen Bedrohung über beide Projektjahre.....	175
<b>Tabelle 10:</b> Lineare Veränderung im wahrgenommenen Verlust über beide Projektjahre .....	176
<b>Tabelle 11:</b> Lineare Veränderung in der wahrgenommenen Binnendifferenzierung über beide Projektjahre .....	178
<b>Tabelle 12:</b> Diskontinuierliche, jahresspezifische Veränderung in der schulischen Selbstwirksamkeitserwartung.....	182
<b>Tabelle 13:</b> Diskontinuierliche, jahresspezifische Veränderung in der Prüfungsängstlichkeit .....	185
<b>Tabelle 14:</b> Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Herausforderung .....	187
<b>Tabelle 15:</b> Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Bedrohung .....	190

<b>Tabelle 16:</b> Diskontinuierliche, jahresspezifische Veränderung im wahrgenommenen Verlust .....	192
<b>Tabelle 17:</b> Diskontinuierliche, jahresspezifische Veränderung in der wahrgenommenen Binnendifferenzierung.....	194
<b>Tabelle 18:</b> Beteiligung der Lehrkräfte an den Workshops .....	196
<b>Tabelle 19:</b> Anzahl an modulspezifischen Beratungssitzungen im Anschluss an die Workshops.....	196
<b>Tabelle 20:</b> Wahrgenommene modulspezifische Implementation seitens der Schüler ....	197
<b>Tabelle 21:</b> Einfluss der wahrgenommenen Implementation auf die schulische Selbstwirksamkeitserwartung.....	202
<b>Tabelle 22:</b> Einfluss der wahrgenommenen Implementation auf die Prüfungsängstlichkeit .....	204
<b>Tabelle 23:</b> Einfluss der wahrgenommenen Implementation auf das schulbezogene Herausforderungserleben .....	207
<b>Tabelle 24:</b> Einfluss der wahrgenommenen Implementation auf das schulbezogene Bedrohungserleben.....	210
<b>Tabelle 25:</b> Einfluss der wahrgenommenen Implementation auf das schulbezogene Verlusterleben .....	213
<b>Tabelle 26:</b> Einfluss der wahrgenommenen Implementation auf die wahrgenommene Binnendifferenzierung.....	216

## Abbildungsverzeichnis

<b>Abbildung 1:</b> Grundmodell der klassischen Motivationspsychologie nach Rheinberg & Vollmeyer (2012), S. 70. ....	10
<b>Abbildung 2:</b> Zeitlicher Verlauf des Projektes.....	96
<b>Abbildung 3:</b> Zielscheibe zur Klimadiagnose.....	99
<b>Abbildung 4:</b> Das Proaktive Handlungsregulationsmodell.....	109
<b>Abbildung 5:</b> Wirkmodell des FOSIS2-Programms .....	112
<b>Abbildung 6:</b> Design zur Evaluation der Wirksamkeit.....	124
<b>Abbildung 7:</b> Design zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation .....	125
<b>Abbildung 8:</b> Ausschnitt aus der ersten modulspezifischen Rückmeldung zum Klassenklima .....	142
<b>Abbildung 9:</b> Screenshot des umstrukturierten SPSS-Datensatzes mit einer kontinuierlichen Zeitvariablen.....	149
<b>Abbildung 10:</b> Screenshot des umstrukturierten SPSS-Datensatzes mit zwei Dummy-Zeitvariablen.....	154
<b>Abbildung 11:</b> Lineare Veränderung in der schulischen Selbstwirksamkeitserwartung über beide Projektjahre (T1-T3 = Messzeitpunkt 1-3) .....	169
<b>Abbildung 12:</b> Lineare Veränderung in der Prüfungsängstlichkeit über beide Projektjahre (T1-T3 = Messzeitpunkt 1-3) .....	171
<b>Abbildung 13:</b> Diskontinuierliche, jahresspezifische Veränderung in der schulischen Selbstwirksamkeitserwartung (T1-T3 = Messzeitpunkt 1-3).....	183
<b>Abbildung 14:</b> Diskontinuierliche, jahresspezifische Veränderung in der Prüfungsängstlichkeit (T1-T3 = Messzeitpunkt 1-3) .....	185
<b>Abbildung 15:</b> Wahrgenommene modulspezifische Implementation seitens der Schüler .....	197





## **Anhang**

## Inhaltsverzeichnis des Anhangs

<b>Anhang A</b>	<b>Ergebnisse der Wave-Nonresponse-Analysen.....</b>	<b>2</b>
A1	Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3) .....	2
A2	Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2).....	6
<b>Anhang B</b>	<b>Skalenhandbuch.....</b>	<b>9</b>
B1	Schulische Selbstwirksamkeitserwartung .....	9
B2	Prüfungsängstlichkeit .....	9
B3	Schulbezogenes Stresserleben – Herausforderung, Bedrohung und Verlust .....	10
B4	Binnendifferenzierung.....	12
B5	Skalen der modulspezifischen Rückmeldungen.....	13
<b>Anhang C</b>	<b>Voraussetzungen für die Anwendung von Mehrebenenanalysen und Vorgehen bei der Mehrebenenmodellierung in dieser Studie.....</b>	<b>15</b>
C1	Voraussetzungen für die Anwendung von Mehrebenenanalysen .....	15
C2	Vorgehen bei der Mehrebenenmodellierung.....	18
<b>Anhang D</b>	<b>Deskriptive Statistiken .....</b>	<b>24</b>
<b>Anhang E</b>	<b>Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation.....</b>	<b>25</b>
<b>Anhang F</b>	<b>Eigenständigkeitserklärung .....</b>	<b>31</b>

## Anhang A Ergebnisse der Wave-Nonresponse-Analysen

### A1 Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3)

**Tabelle A1: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3) – Completer versus Dropout**

	Completer-SP (n = 314)	Dropout-SP (n = 78)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	169(53.8%)	36(46.2%)	0.337
männlich	143(45.6%)	39(50%)	
Keine Angabe	2(0.6%)	3(3.8%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	12;4	12;6	0.107
SD (Monate)	7	9	
<b><i>Nationalität</i></b>			
Deutsch	246(78.3%)	55(70,5%)	0.130
andere Nationalität	47(15.0%)	17(21,8%)	
Keine Angabe	21(6.7%)	6(7.7%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.91(.54)	2.84(.58)	.292
Prüfungsängstlichkeit	3.05(.68)	3.01(.75)	.685
Herausforderungserleben	2.60(.70)	2.57(.77)	.741
Bedrohungserleben	2.20(.60)	2.25(.69)	.500
Verlusterleben	1.62(.70)	1.76(.74)	.130
Binnendifferenzierung	2.05(.70)	2.07(.74)	.862

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

**Tabelle A2: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3) – Completer versus Incompleter T1**

	Completer-SP (n = 314)	Incompleter-SP (n = 100)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	167(53.2%)	49(49%)	.606
männlich	142(45.2%)	47(47%)	
Keine Angabe	5(1.6%)	4(4%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	12;4	12;6	.095
SD (Monate)	8	10	
<b><i>Nationalität</i></b>			
Deutsch	241(76.8%)	68(68%)	.149
andere Nationalität	46(14.6%)	20(20%)	
Keine Angabe	27(8.6%)	12(12%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.91(.54)	2.83(.59)	.180
Prüfungsängstlichkeit	3.05(.68)	3.02(.72)	.723
Herausforderungserleben	2.60(.70)	2.56(.77)	.618
Bedrohungserleben	2.20(.60)	2.23(.67)	.605
Verlusterleben	1.62(.70)	1.75(.71)	.108
Binnendifferenzierung	2.05(.70)	2.06(.72)	.982

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

**Tabelle A3: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3) – Completer versus Incompleter T2**

	Completer-SP (n = 314)	Incompleter-SP (n = 165)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	163(51.9%)	79(47.9%)	.433
männlich	140(44.6%)	80(48.5%)	
Keine Angabe	11(3.5%)	6(3.6%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	13;3	13;6	<b>.000</b>
SD (Monate)	7	9	
<b><i>Nationalität</i></b>			
Deutsch	241(76.7%)	118(71.5%)	.259
andere Nationalität	53(16.9%)	35(21.2%)	
Keine Angabe	20(6.4%)	12(7.3%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.92(.56)	2.77(.61)	<b>.012</b>
Prüfungsängstlichkeit	2.85(.733)	2.79(.74)	.406
Herausforderungserleben	2.30(.72)	2.26(.63)	.474
Bedrohungserleben	2.13(.71)	2.26(.77)	.061
Verlusterleben	1.88(.74)	1.98(.78)	<b>.014</b>
Binnendifferenzierung	2.18(.65)	2.08(.65)	.112

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

**Tabelle A4: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit (T1, T2 und T3) – Completer versus Incompleter T3**

	Completer-SP (n = 314)	Incompleter-SP (n = 156)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	162(51.6%)	76(48.7%)	.623
männlich	141(44.9%)	73(46.8%)	
Keine Angabe	11(3.5%)	7(4.5%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	14;0	14;2	<b>.002</b>
SD (Monate)	6	9	
<b><i>Nationalität</i></b>			
Deutsch	222(70.7%)	115(73.7%)	.898
andere Nationalität	48(15.2%)	24(15.4%)	
Keine Angabe	44(14.1%)	17(10.9%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.96(.54)	2.84(.58)	<b>.032</b>
Prüfungsängstlichkeit	2.84(.73)	2.86(.77)	.724
Herausforderungserleben	2.32(.72)	2.22(.80)	.198
Bedrohungserleben	2.04(.41)	2.16(.79)	.116
Verlusterleben	1.76(.77)	1.89(.81)	.099
Binnendifferenzierung	2.23(.66)	2.08(.61)	<b>.018</b>

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

## A2 Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2)

**Tabelle A5: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2) – Completer versus Dropout**

	Completer-SP (n = 130)	Dropout-SP (n = 27)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	66(50.8%)	8(29.6%)	.062
männlich	64(49.2%)	18(66.7%)	
Keine Angabe	--	1(3.7%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	12;4	12;6	.264
SD (Monate)	8	9	
<b><i>Nationalität</i></b>			
Deutsch	92(70.8%)	16(59.3%)	.233
andere Nationalität	22(16.9%)	7(25.9%)	
Keine Angabe	16(12.3%)	4(14.8%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.92(.51)	3.06(.55)	.210
Prüfungsängstlichkeit	3.02(.68)	3.05(.65)	.857
Herausforderungserleben	2.76(.67)	2.95(.72)	.196
Bedrohungserleben	2.20(.63)	2.11(.72)	.499
Verlusterleben	1.53(.67)	1.69(.70)	.270
Binnendifferenzierung	1.97(.73)	1.89(.73)	.545

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

**Tabelle A6: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2) – Completer versus Incompleter T1**

	Completer-SP (n = 130)	Incompleter-SP (n = 88)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	64(49.2%)	37(42.0%)	.291
männlich	63(48.5%)	49(55.7%)	
Keine Angabe	3(2.3%)	2(2.3%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	12;4	12;4	.837
SD (Monate)	8	9	
<b><i>Nationalität</i></b>			
Deutsch	88(67.6%)	55(62.5%)	.236
andere Nationalität	21(16.2%)	20(22.7%)	
Keine Angabe	21(16.2%)	13(14.8%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.93(.52)	2.92(.55)	.973
Prüfungsängstlichkeit	3.02(.68)	3.01(.67)	.834
Herausforderungserleben	2.76(.67)	2.72(.74)	.673
Bedrohungserleben	2.20(.63)	2.21(.61)	.972
Verlusterleben	1.53(.67)	1.65(.70)	.203
Binnendifferenzierung	2.00(.73)	2.14(.71)	.103

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung



**Tabelle A7: Wave-Nonresponse-Analysen der Daten zur Evaluation der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation (T1, MSR1-3 und T2) – Completer versus Incompleter T2**

	Completer-SP (n = 130)	Incompleter-SP (n = 93)	p für $\chi^2$ -/ T-Test
<b><i>Geschlecht:</i></b>			
weiblich	66(50.8%)	44(47.3%)	.758
männlich	62(47.7%)	45(46.2%)	
Keine Angabe	2(1.5%)	4(4.3%)	
<b><i>Alter</i></b>			
M (Jahre; Monate)	13;2	13;4	.062
SD (Monate)	7	7	
<b><i>Nationalität</i></b>			
Deutsch	88(67.6%)	67(72.1%)	.390
andere Nationalität	30(23.1%)	17(18.2%)	
Keine Angabe	12(9.3%)	9(9.7%)	
<b><i>Kriterien: M(SD)</i></b>			
Schulische Selbstwirksamkeitserwartung	2.89(.52)	2.84(.53)	.485
Prüfungsängstlichkeit	2.65(.73)	2.68(.72)	.743
Herausforderungserleben	2.39(.71)	2.28(.71)	.258
Bedrohungserleben	2.05(.69)	2.09(.67)	.644
Verlusterleben	1.75(.67)	1.81(.71)	.486
Binnendifferenzierung	2.26(.64)	2.38(.68)	.188

*Anmerkungen:* SP = Stichprobe; M = Mittelwert; SD = Standardabweichung

## Anhang B Skalenhandbuch

### B1 Schulische Selbstwirksamkeitserwartung

Die von Jerusalem und Satow (1999) entwickelte Skala *Schulische Selbstwirksamkeitserwartung* erfasst mittels fünf Items die Einschätzung der eigenen Kompetenzen im Umgang mit schulischen Anforderungen.

#### *Items der Skala Schulische Selbstwirksamkeitserwartung*

1. Ich kann auch die schwierigen Aufgaben im Unterricht lösen, wenn ich mich anstrengende.
2. Wenn ich eine schwierige Aufgabe an der Tafel lösen soll, glaube ich, dass ich das schaffen werde.
3. Selbst wenn ich mal längere Zeit krank sein sollte, kann ich immer noch gute Leistungen erzielen.
4. Auch wenn ein Lehrer an meinen Fähigkeiten zweifelt, bin ich mir sicher, dass ich gute Leistungen erzielen kann.
5. Ich bin mir sicher, dass ich auch dann noch meine gewünschten Leistungen erreichen kann, wenn ich mal eine schlechte Note bekommen habe.

**Tabelle A8: Kennwerte der Skala Schulische Selbstwirksamkeitserwartung**

	T1 (n = 409 )	T2 (n = 478)	T3 (n = 470)
M(SD)	2.99(.55)	2.87(.59)	2.91(.55)
Md	3.00	3.00	3.00
Mo	3.00	3.00	3.00
Schiefe	-.37	-.35	-.25
$\alpha$	.72	.75	.77

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

### B2 Prüfungsängstlichkeit

Die in der Dissertation verwendete Skala zur Erfassung der *Prüfungsängstlichkeit* wurde in Anlehnung an die Besorgtheits-Komponente des deutschen TAI von Hodapp et al. (Hodapp, Laux & Spielberger, 1982) entwickelt und bereits mehrfach in Vorgängerprojekten erprobt (siehe z. B. Jerusalem, 2008; Röder, 2008). Mit der aus fünf Items bestehenden

Skala wird die Besorgtheit als Bestandteil der Ängstlichkeit im Sinne einer überdauernden Persönlichkeitseigenschaft erhoben.

### ***Items der Skala Prüfungsängstlichkeit***

1. Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, mache ich mir Sorgen, ob ich auch alles schaffe.
2. Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, frage ich mich, ob meine Leistung ausreicht.
3. Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, denke ich daran, wie wichtig mir ein gutes Ergebnis ist.
4. Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, bin ich besorgt, dass etwas schief laufen könnte.
5. Wenn ich in der Schule eine Arbeit/einen Test schreiben muss, denke ich daran, was passiert, wenn ich schlecht abschneide.

**Tabelle A9: Kennwerte der Skala Prüfungsängstlichkeit**

	T1 (n = 394)	T2 (n = 476)	T3 (n = 469)
M(SD)	3.04(.69)	2.83(.74)	2.85(.74)
Md	3.00	3.00	3.00
Mo	3.00	3.00	3.00
Schiefte	-.68	-.49	-.38
$\alpha$	.82	.85	.87

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

## **B3 Schulbezogenes Stresserleben – Herausforderung, Bedrohung und Verlust**

In Anlehnung an die im transaktionalen Stressmodell von Lazarus und Folkman (1984) operationalisierten Stresseinschätzungen Herausforderung, Bedrohung und Verlust, wird das schulbezogene Stresserleben im FOSIS2-Projekt mittels drei gleichnamigen Skalen erfasst. Die Skalen bestehen jeweils aus drei Items, die ursprünglich von Jerusalem (1990) entwickelt und im Rahmen von FOSIS2-Vorgängerprojekten an den schulischen Kontext angepasst wurden (Jerusalem, 2005, 2008).

**Items der Subskala wahrgenommene Herausforderung**

1. Ich freue mich auf jeden Schultag, weil ständig interessante und neue Dinge auf mich zukommen.
2. Die Schule ist interessant, weil ich täglich neu herausgefordert werde.
3. Ich bin immer gespannt, wie ich die täglichen Anforderungen in der Schule meistern kann.

**Tabelle A10: Kennwerte der Subskala wahrgenommene Herausforderung**

	T1 (n = 411)	T2 (n = 475)	T3 (n = 469)
M(SD)	2.59(.72)	2.30(.70)	2.29(.75)
Md	2.67	2.33	2.33
Mo	2.67	2.00	2.00
Schiefe	.02	.12	.24
$\alpha$	.60	.71	.79

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

**Items der Subskala wahrgenommene Bedrohung**

1. Ich mache mir Sorgen wegen der Probleme, die in der Schule auf mich zukommen.
2. Ich befürchte, dass ich mit meinen Problemen in der Schule nicht zurechtkomme.
3. Ich denke oft daran, dass ich an den schulischen Anforderungen scheitern könnte.

**Tabelle A11: Kennwerte der Subskala wahrgenommene Bedrohung**

	T1 (n = 410)	T2 (n = 476)	T3 (n = 470)
M(SD)	2.21(.62)	2.18(.73)	2.09(.73)
Md	2.00	2.00	2.00
Mo	2.00	2.00	2.00
Schiefe	.22	.35	.38
$\alpha$	.54	.71	.77

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

**Items der Subskala wahrgenommener Verlust**

1. Ich bin niedergeschlagen, weil ich momentan in der Schule einfach nicht zurechtkomme.
2. Meine augenblickliche schulische Lage ist hoffnungslos.
3. Die Schule meint es zurzeit schlecht mit mir.

**Tabelle A12: Kennwerte der Subskala wahrgenommener Verlust**

	T1 (n = 411)	T2 (n = 475)	T3 (n = 470)
M(SD)	1.65(.71)	1.86(.76)	1.80(.79)
Md	1.50	1.67	1.67
Mo	1.00	1.00	1.00
Schiefe	1.057	.75	.78
$\alpha$	.60	.75	.82

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

**B4 Binnendifferenzierung**

Die ebenfalls nur aus drei Items bestehende Skala *Binnendifferenzierung* wurde in Anlehnung an den „Individualized Classroom Environment Questionnaire – ICEQ“ von Fraser (1982) entwickelt. Aus der Perspektive der Schüler erfassen die Items der Skala, inwiefern die Schüler binnendifferenzierende beziehungsweise individualisierende Maßnahmen im Unterricht erleben, das heißt, ob sich die Lehrkräfte sowohl bei den Aufgabenstellungen als auch bei Leistungsrückmeldungen am individuellen Leistungsstand der Schüler orientieren.

**Items der Skala Binnendifferenzierung**

1. Unsere Lehrer geben Schülern im Unterricht unterschiedliche Aufgaben.
2. Die weniger guten Schüler erhalten leichtere Aufgaben als die guten Schüler.
3. Im Unterricht wird von guten Schülern mehr verlangt als von weniger guten Schülern.

**Tabelle A13: Kennwerte der Skala Binnendifferenzierung**

	T1 (n = 400)	T2 (n = 473)	T3 (n = 470)
M(SD)	2.05(.70)	2.15(.65)	2.18(.64)
Md	2.00	2.00	2.17
Mo	2.00	2.00	2.33
Schiefe	.32	.25	.32
$\alpha$	.56	.46	.54

*Anmerkungen:* T1-T3 = Messzeitpunkt 1-3; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „Trifft nicht zu“ (1), „Trifft kaum zu“ (2), „Trifft eher zu“ (3) und „Trifft genau zu“ (4)

## **B5 Skalen der modulspezifischen Rückmeldungen**

Mithilfe der Handreichungen, in denen die Förderstrategien der FOSIS2-Module festgehalten sind, wurden modulspezifische Skalen entwickelt, deren Items die Kernprinzipien der einzelnen FOSIS2-Module erfassen. Da die Items zur Implementation der Module für das Fach Mathematik von allen Interventionsschülern beantwortet wurden, eignet sich dieses Fach am ehesten für die Beurteilung der psychometrischen Qualität der Skalen. Die nachfolgend berichteten Kennwerte basieren entsprechend auf den Skalenwerten des Fachs Mathematik. Für die Berechnungen in der Dissertation wurden die modul- und fachspezifischen Skalenwerte zu einem immer modulspezifischen, fachübergreifenden Implementationsindikator aggregiert.

### ***Implementationscheckitems – Klassenklima***

„Lieber Schülerin, lieber Schüler, wie hast Du den Unterricht **in den letzten Wochen** erlebt?“

1. Wir haben über das Klassenklima gesprochen.
2. Wir haben Klassenregeln erstellt.
3. Gemeinsam mit dem Lehrer haben wir darauf geachtet, dass die Klassenregeln eingehalten werden.
4. Unser Lehrer hat nach unserer Meinung zum Unterricht gefragt.
5. Wenn unser Lehrer nach unserer Meinung gefragt hat, war er auch bereit etwas zu verändern.

**Implementationscheckitems – Transparenz**

„In den letzten Wochen hat uns unser Lehrer vor Tests/Klassenarbeiten ganz genau mitgeteilt...“

1. ...welche Themen und Anforderungen dran kommen
2. ...auf welche Weise wir uns gut vorbereiten können (Mitschriften, Arbeitsblätter, Lehrbuch).
3. ...wie die einzelnen Aufgaben/Teile der Tests/Arbeiten bewertet werden (Punktzahl, Gewichtung der Aufgaben).
4. ...was wir leisten müssen, um eine bestimmte Note zu erreichen.
5. ...welche Art von Aufgaben uns erwartet (Multiple Choice, Textaufgabe).

**Implementationscheckitems – Lern- und Leistungsraum**

„In den letzten Wochen...“

1. ...hatten wir die Möglichkeit, lernen und üben zu können, ohne dass es dafür Noten gab.
2. ...hat uns unser Lehrer immer ganz deutlich gesagt, wann wir lernen und üben konnten, ohne dafür eine Note zu bekommen.
3. ...wenn uns unser Lehrer gesagt hat, dass wir lernen und üben können, ohne uns zu benoten, hat er sich auch daran gehalten.
4. ...hat mein Lehrer bei meinen Leistungen darauf geachtet, ob ich mich verbessere oder verschlechtere.
5. ...hat mir mein Lehrer Mut gemacht, wenn ich mal etwas nicht konnte.

**Tabelle A14: Kennwerte der Skalen der modulspezifischen Rückmeldungen**

	MSR1 ImpS.kk (n = 186)	MSR2 ImpS.tp (n = 196)	MSR13 ImpS.llr (n = 188)
M(SD)	2.08(.74)	2.86(.65)	2.78(.66)
Md	2.00	2.80	2.80
Mo	3.00	2.80	3.00
Schiefe	.36	-.38	-0.40
$\alpha$	.79	.77	.74

*Anmerkungen:* MSR1-3 = Modulspezifische Rückmeldung 1-3; ImpS = Bezeichnung der Variablen für die wahrgenommene Implementation seitens der Schüler; KK = Klassenklima; TP = Transparenz; LLR = Lern- und Leistungsraum; M = Mittelwert; SD = Standardabweichung; Md = Median; Mo = Modus;  $\alpha$  = Cronbachs  $\alpha$ ; Antwortformat = „nie“ (1), „selten“ (2), „oft“ (3) und „bei jeder Gelegenheit“ (4)

## **Anhang C Voraussetzungen für die Anwendung von Mehrebenenanalysen und Vorgehen bei der Mehrebenenmodellierung in dieser Studie**

### **C1 Voraussetzungen für die Anwendung von Mehrebenenanalysen**

Neben der Existenz von Hierarchien in den zu analysierenden Daten müssen wie bei anderen statistischen Verfahren auch zur Durchführung von Mehrebenenanalysen zusätzliche Voraussetzungen beachtet werden. Diese werden nachfolgend beschrieben.

#### ***Substantielle Varianz zwischen den Gruppen und Intraklassen-Korrelationen***

Eine hierarchisch strukturierte, aus Teilgruppen bestehende Stichprobe lässt vermuten, dass sich die Individuen einer Gruppe ähnlicher sind als die Individuen verschiedener Gruppen und dass sich bezüglich der zu untersuchenden Fragestellung Unterschiede zwischen den Gruppen zeigen. Es kann aber sein, dass diese Unterschiede trotz der Existenz solcher Untergruppen oder sogenannter ‚Cluster‘ (z. B. Klassen) nicht existieren beziehungsweise sehr gering sind und somit die für traditionelle statistische Verfahren notwendige Voraussetzung der unabhängigen Beobachtungen eingehalten wird. Ein mehrebenenanalytisches Vorgehen wäre in diesem Fall überflüssig und die Fragestellungen könnten mithilfe klassischer Verfahren (z. B. Varianz- oder Regressionsanalysen) untersucht werden. Gibt es jedoch eine substantielle Varianz zwischen den Gruppen, sollten Mehrebenenanalysen berechnet werden (Garson, 2013c). Gleichmaßen wie in querschnittlichen Analysen dient auch bei Längsschnittuntersuchungen das *Nullmodell* (*Unconditional Means Model*) der Überprüfung dieser Voraussetzung (Singer & Willett, 2003). In diesem Modell werden noch keinerlei Prädiktoren aufgenommen, wodurch die Varianz der Achsenabschnitte (‚intercepts‘) zwischen den Gruppen der jeweiligen Ebene ermittelt wird. Weicht diese signifikant von Null ab, kann von einer substantiellen Variation auf der entsprechenden Ebene ausgegangen und weiterführende Mehrebenenmodelle mit Prädiktoren sollten modelliert werden (Singer & Willett, 2003). Neben der Signifikanztestung der einzelnen Varianzkomponenten können durch die Modellierung des Nullmodells die sogenannten *Intraklassen-Korrelationen* (*Intraclass-Correlations*, ICCs) als weiteres Prüfkriterium für jede Ebene berechnet werden. Die ICC beschreibt den Anteil der Varianz zwischen den Klassen an der Gesamtvarianz und erfasst somit, wie viel Varianz der Ebene noch aufzu-



klären ist (Hox, 2010; Singer & Willett, 2003).<sup>1</sup> Für die Einschätzung und Bewertung der Höhe der Intra-Klassen-Korrelation liegen verschiedene Empfehlungen unterschiedlicher Autoren vor. Hox (2010) empfiehlt eine weitergehende Modellierung ab einer ICC von .05, wobei er diese als klein, eine ICC ab .10 als mittel und eine ICC ab .15 als groß bewertet. Luke (2004) befürwortet bereits eine Mehrebenenmodellierung ab einer ICC größer Null.

Wenn die Varianz zwischen den Klassen in dieser Studie nicht signifikant wird und die ICC sehr klein ausfällt, lässt sich dies im Hinblick auf die Evaluationsfragestellung bereits als ein Zeichen für einen nicht vorhandenen Effekt der Intervention werten. Denn entsprechend den Hypothesen dieser Arbeit werden Unterschiede zwischen den Interventions- und Kontrollklassen erwartet. Eine nicht signifikante Varianz zwischen den Klassen und eine ICC, die gegen Null geht, weisen somit bereits darauf hin, dass sich Interventions- und Kontrollklassen nicht substantiell unterscheiden. Die Modellierung der Daten könnte folglich bereits an dieser Stelle beendet werden. Zur Vereinheitlichung der Methode und des Vorgehens sowie aus Gründen der Vollständigkeit werden jedoch auch für Nullmodelle mit einer sehr kleinen ICC die nachfolgenden Modelle mit Prädiktoren berechnet. Grundsätzlich kann für jede Ebene eine eigene ICC berechnet werden. Da für die Beantwortung der Frage nach der Wirksamkeit des Programms die Variation zwischen den Klassen interessiert, wird hier jeweils die ICC der Klassenebene berichtet.

### ***Stichprobengröße***

Wie bei anderen statistischen Verfahren ist eine angemessene Stichprobengröße auch im Rahmen von Mehrebenenanalysen eine entscheidende Voraussetzung für eine akkurate Schätzung der Parameter und eine gute Teststärke. Die Frage nach der notwendigen Stichprobengröße stellt hierbei eine besondere Herausforderung dar, denn in Mehrebenenmodellen müssen Anforderungen an die Stichprobengröße mehrerer Ebenen gleichzeitig erfüllt werden (Hox, 2010). Nach Hox (2010) ist die Modellierung einer extra Ebene für die Messzeitpunkte ab drei Messzeitpunkten möglich. Sind die festen Effekte von besonderem Interesse empfiehlt er für jedes Level eine Mindestgröße von 30, also 30 Gruppen und 30

---

<sup>A1</sup> In der vorliegenden Arbeit wird die ICC für das Drei-Level-Modell zur Beantwortung der Evaluationsfragestellung mit folgender Formel berechnet:  $ICC_{Klassen} = u_{00j} / (e_{tij} + r_{0ij} + u_{00j})$ .

Für die Beantwortung der zweiten Fragestellung nach der Wirksamkeit in Abhängigkeit von der wahrgenommenen Implementation wird die ICC wie folgt ermittelt:  $ICC_{Klassen} = u_{0j} / (r_{0i} + u_{0j})$ .

Die Formeln sind aus Hox (2010, S. 34) übernommen und wurden an die hier verwendeten Symbole angepasst. Eine genaue Beschreibung der entsprechenden Koeffizienten findet sich unter 6.5.3 und 6.5.4.

Individuen pro Gruppe. Sollen Interaktionseffekte zwischen den Ebenen oder die Zufallseffekte untersucht werden, müsste die Anzahl der Gruppen und die Anzahl der Individuen pro Gruppe noch größer sein, um akkurate Schätzungen der Parameter zu erhalten (Hox, 2010). Aus ökonomischer Perspektive sind die genannten Anforderungen an die Anzahl der Einheiten pro Ebene im Hinblick auf die Kosten bei der Stichprobenziehung in der Praxis oft nicht einzuhalten (Cohen, 1998; Snijders & Bosker, 1993). So finden sich zahlreiche Studien, die Mehrebenenanalysen mit weniger als 30 Einheiten auf der höchsten Gruppenebene durchführen: Souvignier & Mokhlesgerami (2005) modellieren auf der höchsten Ebene z. B.  $n = 17$  Klassen, Magadi und Desta (2011)  $n = 20$  Länder und Gelman (2006) zeigt am Beispiel von zwei Analysen mit acht und sogar drei Schulen auf der jeweils höchsten Gruppenebene, dass unter Einhaltung bestimmter Verteilungsvoraussetzungen sogar mit einer solch kleinen Anzahl an Gruppen Mehrebenenanalysen möglich sind. Grundsätzlich gilt, dass eine angemessene Anzahl der Gruppen für eine optimale Teststärke und eine genaue Parameterschätzung entscheidender ist als die jeweilige Zellengröße (Garson, 2013; Hox, 2010). Deshalb empfehlen andere Autoren eine Mindestgruppengröße von nur zehn (Felson & Reed, 1986) oder gar fünf Personen pro Cluster (Satow, 1999).

Mit 22 Interventions- und Kontrollklassen für die Analysen zur Programmwirksamkeit, zehn Interventionsklassen für die Analysen zum Implementationseinfluss und durchschnittlich 21 Schülern pro Klassen werden auch in dieser Studie Mehrebenenmodelle mit kleinen Stichproben durchgeführt. In Anbetracht dessen sollten die Ergebnisse dieser Studie lediglich als erste Hinweise für mögliche Effekte des FOSIS2-Programms bewertet werden, die eine zukünftige Überprüfung im Rahmen größer angelegter Untersuchungen erfordern.

### ***Homogenitäts- und Verteilungsvoraussetzungen***

Akkurate Schätzungen in Mehrebenenanalysen setzen die Homogenität der Fehlervarianzen voraus (Ditton, 1998; Garson, 2013), wobei nach Raudenbush und Bryk (2002) die Verletzung der Homogenitätsannahme kein schwerwiegendes Problem darstellt. Heterogenität kann jedoch ein Zeichen für eine mangelhafte Modellspezifikation sein (Garson, 2013). Die meisten Programme stellen Verfahren zur Überprüfung der Homogenität zur Verfügung (Ditton, 1998). Die Software HLM 6.08 bietet den sogenannten ‚*Test of homogeneity of level-1 variance*‘ an, einen Chi-Quadrat-Test, der die Fehlervarianz auf Homogenität (Nullhypothese) prüft. Wird dieser signifikant, ist die Homogenitätsannahme verletzt. Die Ergebnisse des Tests zeigen, dass die Homogenitätsannahme bei den geprüften Fehlervarianzen der hier berechneten Modelle nicht verletzt wird.

Die Prädiktoren und Residuen sollten möglichst normalverteilt sein (Garson, 2013; Greenberg & Phillips, 2013), wobei die Schätzung der festen Effekte gegenüber Verletzungen der Normalverteilungsannahme relativ robust ist (Eid, Gollwitzer & Schmitt, 2013).

## **C2 Vorgehen bei der Mehrebenenmodellierung**

Das Verfahren der Mehrebenenanalyse bietet zahlreiche Möglichkeiten der Modellbildung und verschiedene Kriterien zur Beurteilung der Modellgüte an. Zudem sind in der Regel Entscheidungen über Schätzverfahren der Modellparameter und mögliche Zentrierungen sowie Umkodierungen der Variablen notwendig, um die Ergebnisse sinnvoll interpretieren zu können. Das Vorgehen bei der Modellentwicklung und die Wahl der Gütekriterien, der Schätzverfahren sowie der Zentrierung beziehungsweise Umkodierung werden nachfolgend erläutert.

### ***Modellentwicklung***

Das Vorgehen bei der Modellentwicklung im Rahmen von Mehrebenenanalysen ähnelt dem der multiplen Regression. Die Modellbildung sollte theoretisch und inhaltlich begründet erfolgen (Garson, 2013), alle wesentlichen Prädiktoren einschließen und gleichzeitig so sparsam wie möglich gestaltet werden (Hox, 2010). Singer und Willett (2003) fassen die bei der Auswahl der Prädiktoren zu beachtenden Aspekte folgendermaßen zusammen: „We suggest that you base decisions to enter, retain, and remove predictors on a combination of logic, theory, and prior research, supplemented by judicious hypothesis testing and comparison of model fit“ (S. 105). Wie in multiplen Regressionen ist es möglich, einzelne oder Gruppen von Prädiktoren aufzunehmen (Singer & Willett, 2003) und bei der Modellbildung eine Top-down- oder Bottom-up-Strategie zu verfolgen (Hox, 2010). Das Top-Down-Vorgehen hat in Mehrebenenanalysen den Nachteil, dass gleich zu Beginn sehr komplexe Modelle mit vielen Prädiktoren modelliert werden. Dabei entstehen nicht selten Konvergenzschwierigkeiten und oft sind zahlreiche, langwierige Iterationsprozesse für die Schätzung der Modellparameter erforderlich (Hox, 2010). Daher wird üblicherweise die Bottom-up-Strategie zur Modellbildung gewählt, wobei die Aufnahme der Prädiktoren sukzessive erfolgt. Im Modell verbleiben dann die Prädiktoren, die statistisch signifikant werden und zu einer Verbesserung der Modellgüte beitragen (Dedrick et al., 2009). Eine besondere Herausforderung der Mehrebenenanalyse stellt die Wahl der zu modellierenden Zufallseffekte dar. Je mehr Prädiktoren ein Modell umfasst, desto mehr Entscheidungen müssen darüber getroffen werden, welche Effekte variieren dürfen und welche nicht. Für einige

Effekte lassen sich aus der Theorie Hypothesen darüber ableiten, ob sie variieren dürfen, für andere bestehen Unsicherheiten hinsichtlich einer potentiellen Variation. Neben inhaltlichen Überlegungen geben auch hier die Kriterien zur Beurteilung der Modellgüte (siehe unten) Auskunft darüber, welche Zufallseffekte fixiert werden sollten (Dedrick et al., 2009).

Hox (2010) empfiehlt fünf Schritte zur Modellbildung nach der Bottom-up-Methode: erstens die Modellierung eines Nullmodells ohne jegliche Prädiktoren, zweitens die sukzessive Aufnahme aller möglichen Level-1-Prädiktoren und drittens die sukzessive Aufnahme aller Prädiktoren höherer Ebenen. Dabei verbleiben nur die Prädiktoren im Modell, die signifikant werden und die Kriterien zur Beurteilung der Modellgüte verbessern. Aufgrund inhaltlicher und theoretischer Überlegungen sollen an vierter Stelle Zufallseffekte modelliert werden. Wiederum erfolgt die Entscheidung darüber, welche Zufallseffekte im Modell als variierend belassen werden, auf der Basis der Signifikanz der Parameter und der Kriterien zur Bewertung der Modellgüte. Als letzten Schritt empfiehlt Hox (2010) die Modellierung von Cross-Level-Interaktionen.

Bei der Modellbildung in dieser Studie wurde das Bottom-Up-Vorgehen nach Hox (2010) verwendet. Eine ausführliche Beschreibung aller überprüften Modelle mit und ohne Prädiktoren und der sukzessiven Freistellung einzelner Zufallseffekte würde den Rahmen dieser Arbeit sprengen und die Lesbarkeit erschweren. Singer und Willet (2003) empfehlen deshalb, nicht jedes getestete Modell und jeden Zwischenschritt der Modellbildung zu erläutern, sondern für die Beschreibung des Vorgehens und die Präsentation der Ergebnisse in Publikationen solche Modelle auszuwählen und zu berichten, die die Geschichte der Modellentwicklung nachbilden und wichtige Schritte markieren. Bei der Modellierung von Veränderung bedeutet dies nach den Autoren, zumindest das Nullmodell, das Veränderungsmodell und das finale Zielmodell zur Beantwortung der Fragestellung zu präsentieren, wobei ein Modell nur so lange als final anzusehen ist, bis es durch ein besseres Modell abgelöst wird. In der vorliegenden Arbeit werden bei der Beschreibung des Vorgehens entsprechend das Nullmodell sowie die Veränderungsmodelle – also das lineare, diskontinuierliche und autoregressive Modell (siehe 6.5.3 und 6.5.4) – und schließlich die Zielmodelle zur Beantwortung der Fragestellungen dargestellt. Nur für die Frage nach der Wirksamkeit in Abhängigkeit von der Implementation, existieren durch modulspezifische Implementationsindikatoren verschiedene, miteinander konkurrierende Zielmodelle. Hierbei

werden solche Prädiktoren beibehalten und berichtet, die signifikant ausfallen und die Modellgüte verbessern.

### ***Kriterien zur Beurteilung der Modellgüte***

Für die Beantwortung der Forschungsfrage ist entscheidend, welches der berechneten Mehrebenenmodelle am besten zu den Daten passt. In multiplen Regressionsanalysen wird üblicherweise das Bestimmtheitsmaß, der sogenannte Determinationskoeffizient  $R^2$ , als ein „standardisiertes Maß für die Güte der Vorhersage“ (Eid et al., 2013, S. 569), berichtet. Dieses gibt den Anteil der erklärten Varianz an der Gesamtvarianz an. In Mehrebenenanalysen ist die Bestimmung der Varianzaufklärung komplexer: Da auf mehreren Ebenen gleichzeitig unerklärte Varianzen vorliegen, kann kein einheitlicher Determinationskoeffizient als Gütekriterium berechnet werden (Garson, 2013). Aus dem Vergleich zweier Mehrebenenmodelle ist jedoch die Berechnung eines eigenen sogenannten Pseudo- $R^2$  für jede einzelne Varianzkomponente möglich (Eid et al., 2013; Singer & Willett, 2003).<sup>2</sup>

Eine solche Art der Bestimmung der Varianzaufklärung ist mitunter problematisch (Garson, 2013). Denn anders als der Determinationskoeffizient kann das Pseudo- $R^2$  in Mehrebenenanalysen unter bestimmten Umständen negativ werden. In herkömmlichen Regressionsanalysen führt die zusätzliche Aufnahme eines Prädiktors im Vergleich zum Regressionsmodell ohne diesen Prädiktor in der Regel zu einer geringeren Residualvarianz und zu einem höheren  $R^2$ . Im Gegensatz dazu kann es in Mehrebenenanalysen jedoch vorkommen, dass sich durch die Aufnahme eines Prädiktors im Vergleich zum Modell ohne diesen Prädiktor einzelne Residualvarianzen erhöhen und dadurch negative Pseudo- $R^2$ -Statistiken entstehen. Eine ausführliche Beschreibung dieses Phänomens findet sich bei Snijders und Bosker (1994). Die Autoren erklären, dass unter bestimmten Umständen ein negatives Pseudo- $R^2$  unausweichlich ist und schlagen eine Korrekturformel für solche Fälle vor. Nach Eid, Gollwitzer und Schmitt (2013) kann jedoch auch diese nicht immer ein negatives Pseudo- $R^2$  verhindern. Auch Hox (2010) konstatiert nach einer ausführlichen

<sup>A2</sup> In Mehrebenenanalysen ist das  $R^2$  ein *indirekter* Schätzer der durch die Prädiktoren aufgeklärten Varianz und wird deshalb mit dem Begriff „Pseudo“ versehen (Eid et al., 2013). Das Pseudo- $R^2$  einer Varianzkomponente berechnet sich übertragen in eine für alle Varianzkomponenten gültige Formel wie folgt (z. B. Singer & Willett, 2003, S. 104; Maerten-Rivera, 2013, S. 271):

$$\text{Pseudo}R^2 = \frac{\text{Varianzkomponente (niedrigeres Modell)} - \text{Varianzkomponente (höheres, benachbartes Modell)}}{\text{Varianzkomponente (niedrigeres Modell)}}$$

Dabei ist zu beachten, dass immer der gleiche Varianzparameter ein und derselben Ebene zweier Modelle in die Berechnungen einfließt. Manche Autoren ziehen für die Ermittlung des Pseudo- $R^2$  als Ausgangsmodell grundsätzlich das Nullmodell heran (vgl. z. B. Hox, 2010; Maerten-Rivera, 2013). Eid, Gollwitzer und Schmitt (2013) beschreiben, dass auch benachbarte Modelle mit und ohne den jeweiligen Prädiktor herangezogen werden können. Diese Variante wird in der vorliegenden Studie angewandt.

Beschreibung der aktuellen Debatte über den Umgang mit einem negativen Pseudo- $R^2$ : „there is no current solution for this“ (S. 77). Manche Autoren empfehlen deshalb, gar keine Maße zur Varianzaufklärung zu berichten (Garson, 2013). Grundsätzlich ist bei der Interpretation der Pseudo- $R^2$ -Statistiken Vorsicht geboten und negative Pseudo- $R^2$ -Koeffizienten sollten nicht interpretiert werden (Singer & Willett, 2003). In der vorliegenden Studie werden die Pseudo- $R^2$  für die Varianzkomponenten derjenigen Ebenen ermittelt, auf der Prädiktoren zur Beantwortung der Fragestellungen aufgenommen werden. Negative Pseudo- $R^2$  werden entsprechend der beschriebenen Empfehlungen nicht interpretiert.

Neben dem Pseudo- $R^2$  existieren für Mehrebenenanalysen zusätzliche Möglichkeiten, die Modellgüte zu beurteilen. Beispielsweise kann der Vergleich der sogenannten *Devianz* zweier ineinander geschachtelter Modelle bei der Bewertung eines Modells helfen. Die Devianz wird aus der Likelihood-Funktion abgeleitet, auf der die Schätzverfahren basieren (Eid et al., 2013).<sup>3</sup> Mit einem Chi-Quadrat-Test kann die Differenz der Devianzen zweier geschachtelter Modelle auf Signifikanz getestet werden (Garson, 2013). Nach Eid et al. (2013) sind „zwei Modelle genestet (d.h. ineinander verschachtelt), wenn das eine Modell durch Restriktion beziehungsweise Freisetzung von Modellparametern in das andere Modell überführt werden kann“ (S. 715). Wird die Differenz der Devianzen signifikant, erklärt das Modell mit der kleineren Devianz signifikant besser die Daten. Alle statistischen Programme zur Durchführung von Mehrebenenanalysen bieten den Chi-Quadrat-Test zum Vergleich verschachtelter Modelle an (Garson, 2013).

Zusätzlich zum Chi-Quadrat-Test eignen sich auch der *Akaike-Index* („*Akaike Information Criterion*“, *AIC*)<sup>4</sup> und der *Schwarz-Index* („*Bayesian Information Criterion*“, *BIC*)<sup>5</sup> zur Überprüfung der Modellgüte (Eid et al., 2013; Hox, 2010). Beide Indizes können auch für den Vergleich nicht geschachtelter Modelle herangezogen werden, wobei wie beim Devianztest das Modell mit dem kleineren AIC oder BIC die Daten besser erklärt. Im Unterschied zum AIC fließt in die Berechnung des BIC auch die Stichprobengröße mit ein. Dabei besteht jedoch Unklarheit darüber, welche Stichprobengröße welcher Ebene gemeint ist. Häufig wird das N der höchsten Ebene herangezogen, eine einheitliche Anwendung

<sup>A3</sup> Nach Eid et al. (2013, S. 715) berechnet sich die Devianz aus dem natürlichen Logarithmus der maximierten Likelihood-Funktion multipliziert mit minus Zwei ( $Dev = -2 \cdot \ln(L)$ ).

<sup>A4</sup> Die Formel für den AIC lautet mit „Dev“ für Devianz und „q“ für die Anzahl der zu schätzenden Modellparameter:  $AIC = Dev + 2 \cdot q$  (Eid et al., 2013, S. 716).

<sup>A5</sup> Der BIC berechnet sich mit „Dev“ für Devianz, „q“ für die Anzahl der zu schätzenden Modellparameter und „n“ für die Stichprobengröße wie folgt:  $BIC = Dev + q \cdot \ln(n)$  (Eid et al., 2013, S. 716).

gibt es jedoch nicht (Hox, 2010; Singer & Willett, 2003). Hox (2010) empfiehlt sowohl bei komplexen Modellen mit einer hohen Anzahl an zu schätzenden Parametern sowie bei Modellen mit unterschiedlichen Stichprobengrößen der einzelnen Ebenen die Anwendung des AIC.

Da die Modelle dieser Studie ineinander verschachtelt sind, eine hohe Anzahl an Parametern geschätzt werden müssen und die Stichprobengröße der einzelnen Ebenen unterschiedlich ausfällt, werden zur Überprüfung der Modellgüte neben dem Pseudo- $R^2$  der Chi-Quadrat-Test für den Devianzvergleich sowie der AIC herangezogen.

### ***Schätzverfahren***

Mehrebenenanalysen können mit unterschiedlichen statistischen Softwares durchgeführt werden, die jeweils verschiedene Verfahren zur Schätzung der Parameter anbieten. Das in dieser Studie verwendete Programm HLM 6.08 stellt als Voreinstellung das Schätzverfahren *„Restricted maximum likelihood“* (RML) zur Verfügung. Alternativ kann der Nutzer auch die Methode *„Full maximum likelihood“* (FML) auswählen (Garson, 2013). Obwohl Hox (2010) für kleine Stichproben die Anwendung des RML als Schätzverfahren empfiehlt, wird für die in dieser Arbeit zu schätzenden Modelle die FML-Methode ausgewählt. Denn das RML-Verfahren erlaubt keinen Vergleich von Modellen, die sich in ihren festen Effekten unterscheiden und ineinander geschachtelt sind (Garson, 2013; Greenberg & Phillips, 2013). Genau dies ist jedoch für die hier zu analysierenden Modelle der Fall: Sie sind ineinander geschachtelt und unterscheiden sich in ihren festen Effekten und gerade der Modellvergleich bezüglich der festen Effekte ist von besonderem Forschungsinteresse. Deshalb wird die FML-Methode zur Schätzung der Parameter verwendet.

### ***Zentrierung***

Bei den meisten Mehrebenenmodellen sind Maßnahmen der Zentrierung ratsam, um die Koeffizienten sinnvoll interpretieren zu können. Die sogenannte *Zentrierung um den Gesamtmittelwert* (*„grand mean centering“*) ist das am häufigsten verwendete Verfahren. Eine weitere Möglichkeit stellt die *Zentrierung um den Gruppenmittelwert* (*„group mean centering“*) dar (Garson, 2013). Werden Modelle ohne Zufallskoeffizienten berechnet, ist es irrelevant, ob eine Zentrierung um den Gesamtmittelwert, den Gruppenmittelwert oder gar keine Zentrierung erfolgt. Bei der Modellierung von Zufallseffekten hingegen ändert sich das Modell durch die Zentrierung um die Gruppenmittelwerte. Diese Form der Zentrierung eignet sich besonders gut, wenn für die Beantwortung der Forschungsfrage nicht die zufäl-

lige Variation zwischen den Gruppen interessiert. Denn beim Group-Mean-Centering wird die Variation zwischen den Gruppen gleich ‚Null‘ gesetzt und somit aus dem Modell herausgerechnet (Eid et al., 2013).<sup>6</sup>

In Studien mit einer eigenen Ebene der Zeit und entsprechenden Zeitvariablen ist eine Zentrierung um den Gesamt- oder Gruppenmittelwert nicht üblich (Greenberg & Phillips, 2013). Stattdessen sind Umkodierungen der Zeitvariablen erforderlich, die eine sinnvolle Interpretation der Zeiteinheiten erlauben. In der Regel wird der Startpunkt der Befragungen mit Null kodiert, sodass die Ausgangswerte als Referenzwerte herangezogen werden. Je nach Ausprägung der Zeitvariablen und Fragestellungen können aber auch andere Zeitpunkte als Referenzkategorie gewählt werden (Garson, 2013).

Zur Beantwortung der Frage nach der Wirksamkeit der FOSIS2-Maßnahmen wird eine Kodierung der Zeitvariablen vorgenommen, durch die der erste Messzeitpunkt zum Referenzwert wird. Eine genaue Beschreibung der Kodierung findet sich unter 6.5.3. Für die Analysen zum Einfluss der Implementation liegt keine Zeitebene vor, da diese beispielhaft für das erste Projektjahr überprüft werden und somit nur zwei Messzeitpunkte in die Berechnungen einbezogen werden. Die Wirksamkeit der Intervention soll in Abhängigkeit von der wahrgenommenen Implementation seitens der Schüler untersucht werden. Die Implementation wird dabei auf der Schülerebene modelliert, sodass weder feste Effekte noch Zufallseffekte auf der zweiten Ebene von Interesse sind. Entsprechend wird eine Zentrierung um den Gruppenmittelwert vorgenommen, um dadurch die Unterschiede zwischen den Gruppen zu nivellieren. Eine genauere Beschreibung der Modellbildung und des Vorgehens findet sich unter 6.5.4.

---

<sup>A6</sup> Eine ausführliche Beschreibung und Begründung der Maßnahmen zur Zentrierung geben Enders und Tofighi (2007).



## Anhang D Deskriptive Statistiken

**Tabelle A15: Mittelwerte und Standardabweichungen der Kriterien für Interventions- und Kontrollgruppe**

	IG			KG		
	T1	T2	T3	T1	T2	T3
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Schulische Selbstwirksamkeitserwartung	2.93 (.53)	2.87 (.52)	2.95 (.49)	2.86 (.57)	2.88 (.64)	2.89 (.60)
Prüfungsängstlichkeit	3.02 (.68)	2.66 (.72)	2.74 (.76)	3.07 (.71)	2.98 (.72)	2.94 (.72)
Herausforderungserleben	2.75 (.70)	2.34 (.71)	2.32 (.77)	2.42 (.70)	2.25 (.68)	2.26 (.73)
Bedrohungserleben	2.20 (.62)	2.06 (.68)	1.92 (.76)	2.22 (.62)	2.27 (.76)	2.22 (.74)
Verlusterleben	1.58 (.68)	1.77 (.69)	1.72 (.79)	1.73 (.73)	1.94 (.81)	1.86 (.79)
Binnendifferenzierung	2.04 (.72)	2.31 (.66)	2.26 (.67)	2.07 (.69)	2.01 (.61)	2.12 (.62)

*Anmerkungen:* IG = Interventionsgruppe; KG = Kontrollgruppe; T = Messzeitpunkt; M = Mittelwert; SD = Standardabweichung

## Anhang E Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation

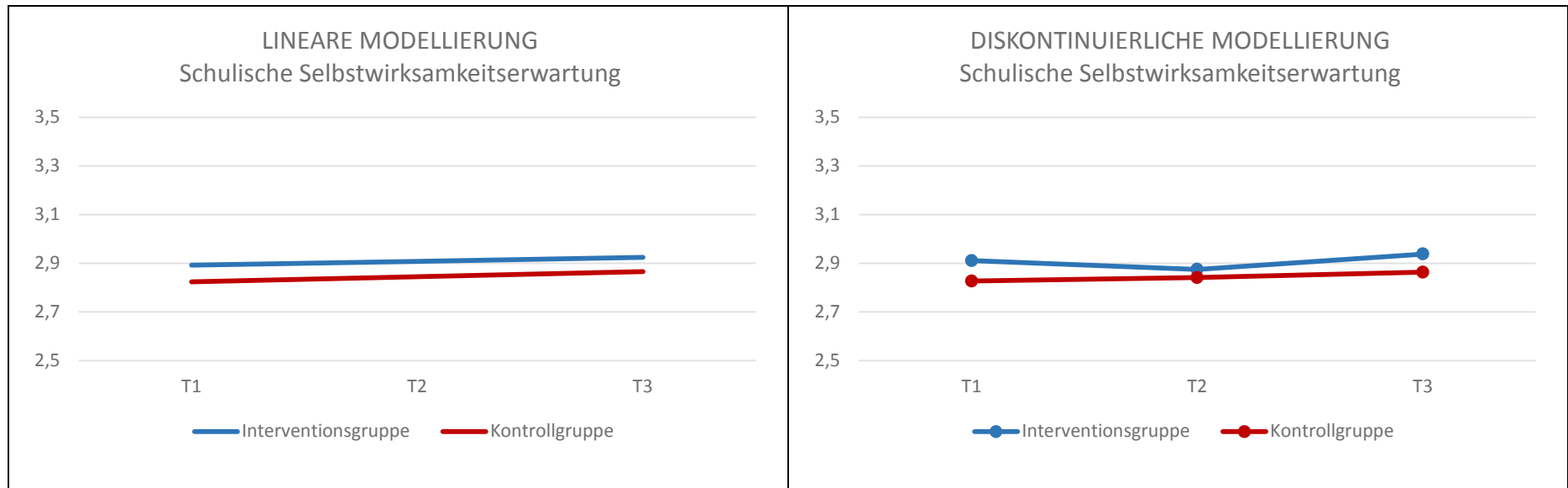


Abbildung A1. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Schulische Selbstwirksamkeitserwartung

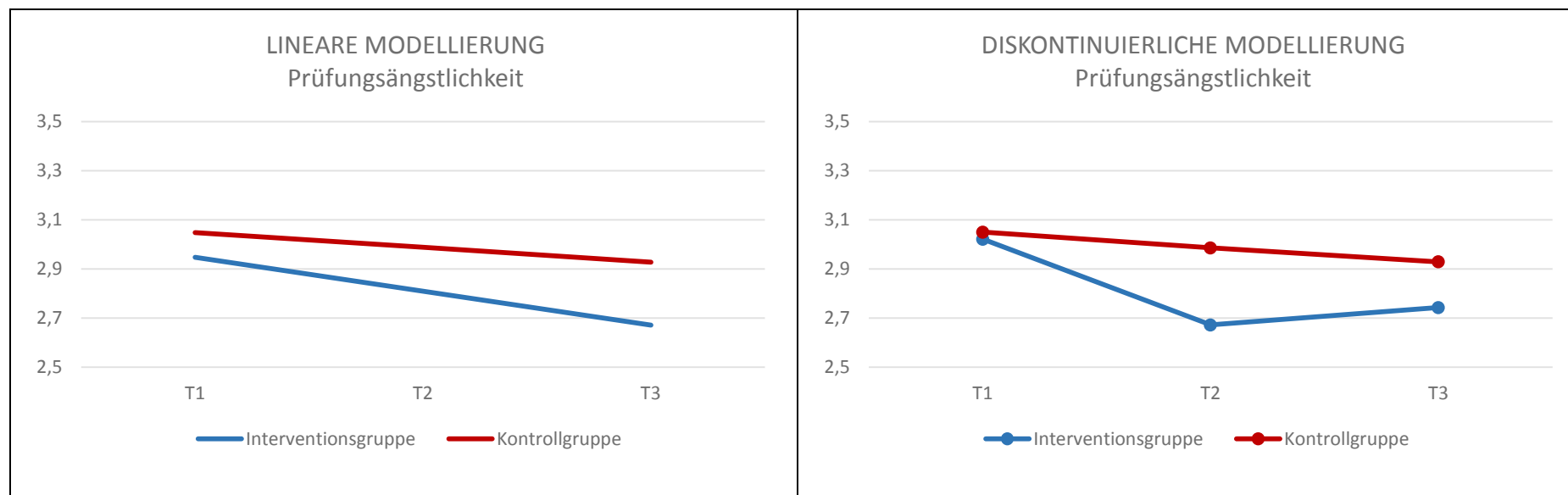


Abbildung A2. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Prüfungsängstlichkeit

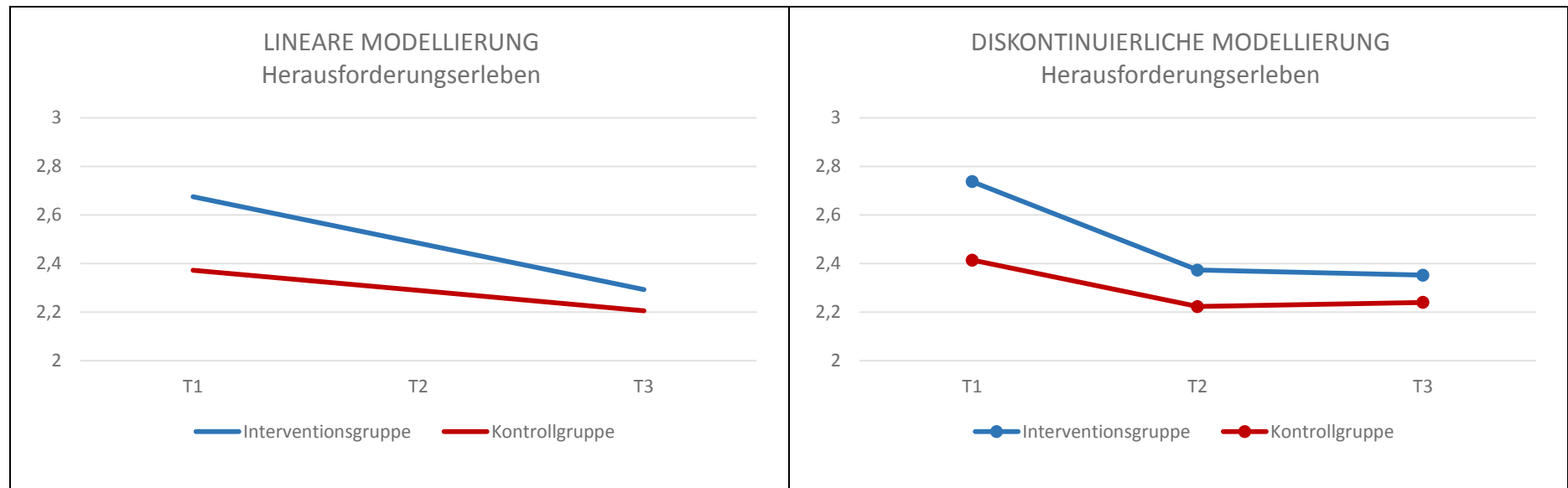
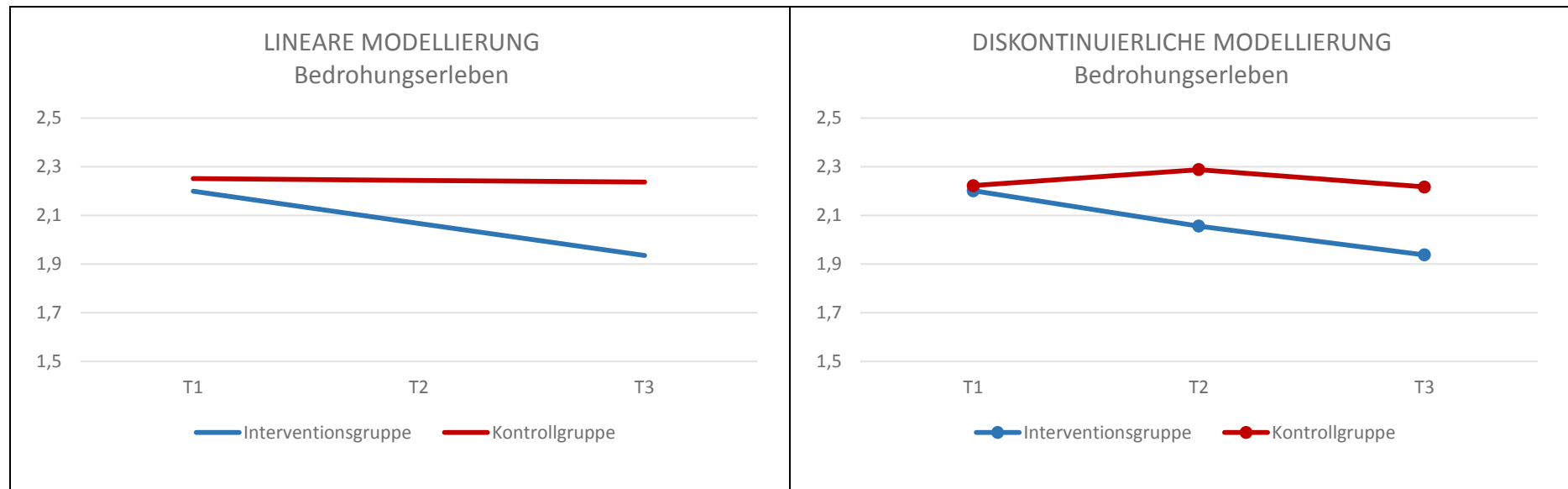


Abbildung A3. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Herausforderungserleben



**Abbildung A4. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Bedrohungserleben**

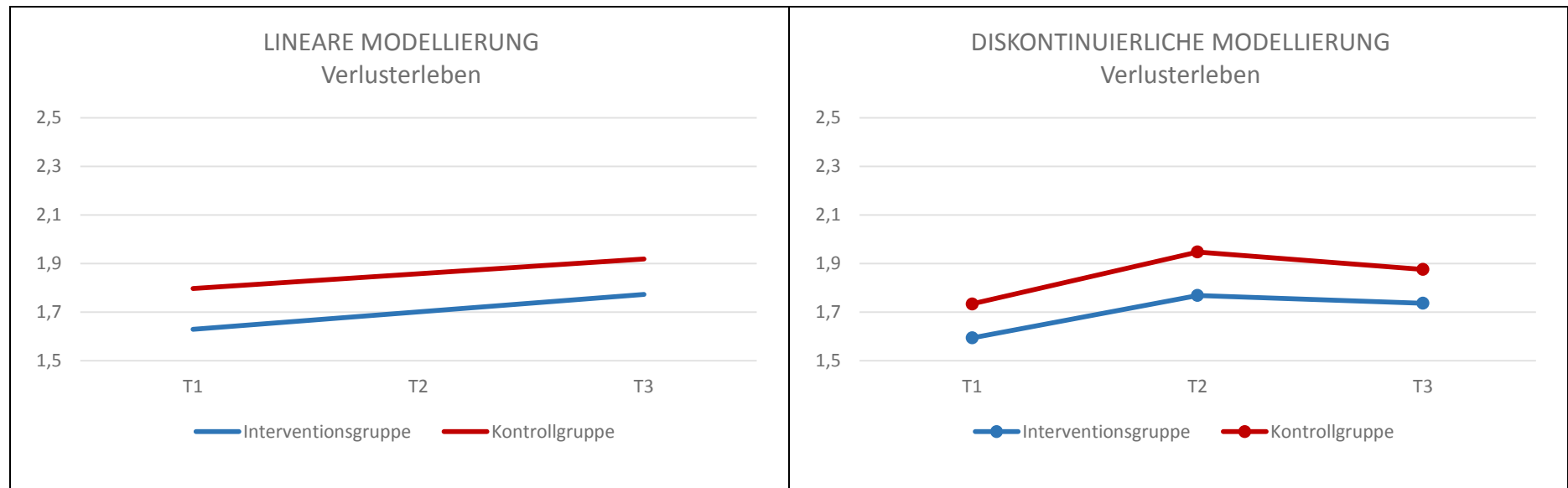
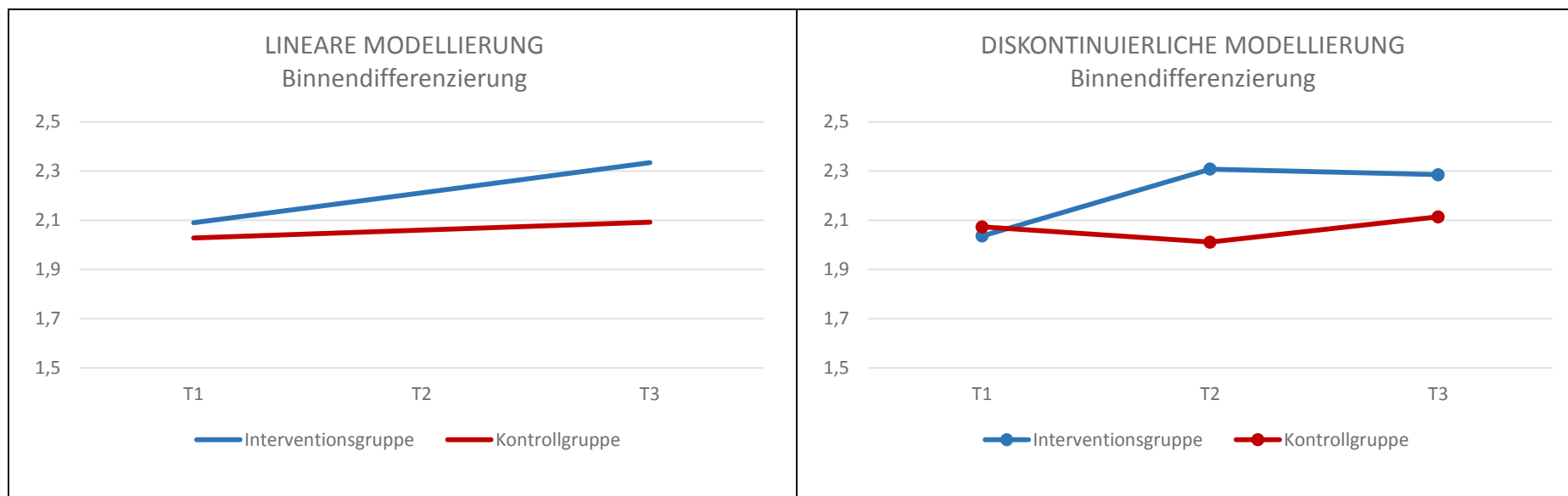


Abbildung A5. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Verlusterleben



**Abbildung A6. Grafische Darstellung der Ergebnisse der Wirksamkeitsevaluation: Binnendifferenzierung**

## **Anhang F   Eigenständigkeitserklärung**

Hiermit bestätige ich, dass ich die vorliegende Arbeit in allen Teilen selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Berlin, den 24. Februar 2015

Katharina Chwallek